



UNIVERSIDADE
LUSÓFONA

CENTRO UNIVERSITÁRIO DO PORTO

FACULDADE DE CIÊNCIAS ECONÓMICAS, SOCIAIS E DA EMPRESA

MESTRADO EM GESTÃO

APLICAÇÃO DE *DATA MINING* À ANÁLISE DA ATIVIDADE DE UMA LAVANDARIA INDUSTRIAL

Dissertação apresentada a provas públicas para a obtenção do grau de Mestre em Gestão, orientada pela Professora Doutora Maria Teresa Ribeiro Candeias e pelo Professor Doutor Hugo Alonso Vilares Monteiro.

João Luís da Costa Ferraz Soares

N.º A22214756

Novembro de 2025

www.ulusofona.pt

Universidade Lusófona
Centro Universitário do Porto
Faculdade de Ciências Económicas, Sociais e da Empresa
Mestrado em Gestão

APLICAÇÃO DE DATA MINING À ANÁLISE DA ATIVIDADE DE UMA LAVANDARIA INDUSTRIAL

João Luís da Costa Ferraz Soares
N.º A22214756

VERSÃO FINAL

Dissertação defendida por João Luís da Costa Ferraz Soares em provas públicas na Universidade Lusófona, Centro Universitário do Porto, no dia 26 de setembro de 2025, perante o júri, nomeado pelo Despacho de Nomeação n.º 164/2025, com a seguinte composição:

Presidente: Prof.^a Doutora Gina Maria Marques de Carvalho Santos, Professora Auxiliar da Universidade Lusófona;

Arguente: Prof. Doutor João Paulo Marôco Domingos, Professor Catedrático da Universidade Lusófona;

Vogal: Prof.^a Doutora Diana da Silva Dias, Professora Catedrática da Universidade Lusófona;

Vogal: Prof.^a Doutora Catarina Joana Vieira Gomes, Professora Associada da Universidade Lusófona.

Orientador: Prof.^a Doutora Maria Teresa Ribeiro Candeias, Professora Auxiliar da Universidade Lusófona.

Este trabalho também foi orientado por Prof. Doutor Hugo Alonso Vilares Monteiro.

Agradecimentos

Em primeiro lugar, quero expressar a minha profunda gratidão aos meus orientadores, Professora Doutora Teresa Candeias e Professor Doutor Hugo Alonso, pela sua orientação, paciência e incentivo ao longo deste percurso. O seu apoio foi essencial para a realização deste trabalho.

Agradeço também à lavandaria Precioso & Morais, Lda. que gentilmente permitiu a utilização dos seus dados para este estudo, sem os quais esta dissertação não seria possível.

Quero ainda dedicar um especial agradecimento aos meus colegas e amigos, que me acompanharam e motivaram em momentos desafiantes, oferecendo palavras de encorajamento e partilhando valiosos conselhos.

À minha família, por estar sempre ao meu lado.

Por fim, agradeço a todas as pessoas que, direta ou indiretamente, contribuíram para a conclusão deste trabalho.

Resumo

A crescente digitalização dos processos industriais gera grandes volumes de dados, criando oportunidades para otimização e tomada de decisões mais precisas. Esta dissertação explora a aplicação de técnicas de *data mining* para melhorar os sistemas de apoio à decisão numa lavandaria industrial. A empresa recolhe dados operacionais, mas ainda não implementou processos para analisá-los ou extrair conhecimento útil. Através da aplicação de algoritmos de *clustering* e modelos preditivos, este estudo visa contribuir para o início da transformação da lavandaria numa organização orientada por dados.

O objetivo principal é desenvolver modelos que auxiliem na tomada de decisões e na previsão, contribuindo para melhorias operacionais. As técnicas de *clustering* são utilizadas para agrupar os dados, identificando padrões nas transações com clientes, tipos de serviços e tratamento de produtos ao longo de um período de seis anos. Esta análise revela tendências que apoiam decisões empresariais relacionadas com a alocação de recursos, segmentação de clientes e estratégias de *marketing*. Adicionalmente, modelos de previsão de séries temporais são aplicados para prever receitas futuras, permitindo que a empresa antecipe a procura e se prepare melhor para as necessidades operacionais futuras.

Palavras-chave: Empresa orientada por dados; *Data mining*; *Clustering*; Lavandaria industrial; Modelos preditivos; Previsão de séries temporais.

Abstract

The increasing digitalization of industrial processes generates vast amounts of data, creating opportunities for optimization and more precise decision-making. This dissertation explores the application of data mining techniques to improve decision support systems within an industrial laundry company. The company collects operational data but has not yet implemented processes to analyse or derive insights from it. By applying clustering algorithms and predictive models, this study aims to contribute to the beginning of the transformation of the laundry into a data-driven organization.

The primary goal is to develop models that assist in decision-making and forecasting, contributing to operational improvements. Clustering techniques are applied to group data, identifying patterns in customer transactions, types of services, and product handling over a six-year period. This analysis reveals trends that inform business decisions related to resource allocation, customer segmentation, and marketing strategies. Additionally, time series forecasting models are applied to predict future revenues, allowing the company to anticipate demand and better prepare for future operational needs.

Keywords: Data-driven company; Data mining; Clustering; Industrial laundry; Predictive models; Time series forecasting

Abreviaturas

DM - *Data mining*

DTW - Dynamic Time Warping

IA – Inteligência artificial

MAE - Mean Absolute Error

MAPE - Mean Absolute Percentage Error

ARIMA - Autoregressive Integrated Moving Average

K-médias - Algoritmo de K-médias

K-Medóides – Algoritmo de K-Medóides

PME – Pequenas e médias empresas

Índice

1.	Introdução.....	2
1.1.	Aplicação de <i>data mining</i> à análise da atividade de uma lavandaria industrial	5
1.2.	Organização da dissertação.....	8
2.	Revisão de literatura	10
2.1.	Empresa orientada por dados	12
	Integração de dados.....	13
	Processamento de dados.....	14
	Armazenamento de dados	14
	Modelo operacional de dados	14
	Chief data officer	14
	Ecosistema de dados	15
	Gestão de dados.....	15
2.2.	<i>Data mining</i>	16
3.	Métodos de <i>clustering</i> e de previsão.....	24
3.1.	<i>Clustering</i>	25
3.1.1.	Algoritmos de <i>Clustering</i>	27
3.1.2.	Coeficiente silhueta.....	33
3.2.	Previsão de séries temporais	34
3.2.1.	Padrões de séries temporais	37
3.2.2.	Métodos de previsão.....	39
4.	Lavandaria Industrial.....	46
5.	Metodologia	52
5.1.	Recolha e processamento de dados	53
5.2.	<i>Clustering</i>	57
5.3.	Previsão de séries temporais	58
5.3.1.	Métodos simples.....	59
5.3.2.	Métodos regressivos e autoregressivo.....	59
5.3.3.	Análise de resíduos	60

6.	Resultados e discussão	62
6.1.	Análise primária de dados	63
6.1.1.	Total líquido por cliente	63
6.1.2.	Total líquido por tipo de lavagem.....	65
6.1.3.	Total líquido por tipo de peça	67
6.2.	<i>Clustering</i>	68
6.2.1.	Aplicação dos algoritmos	69
6.2.2.	Resumo dos resultados ótimos de <i>clustering</i>	74
6.2.3.	Análise dos <i>clusters</i>	77
6.3.	Previsão de séries temporais	94
6.3.1.	Total líquido global	95
6.3.2.	Total líquido por cliente	101
6.3.3.	Total líquido por tipo de lavagem.....	103
6.3.4.	Total líquido por tipo de peça	110
6.3.5.	Análise global dos resultados	112
7.	Conclusões	116
7.1.	Resultados	120
7.2.	Relação com outros estudos	128
7.3.	Trabalho futuro	129
8.	Bibliografia	132
9.	Apêndices	138
9.1.	Resultados <i>Clustering</i>	139
9.1.1.	Total líquido por Cliente	139
9.1.2.	Total líquido por Tipo de Lavagem	142
9.1.3.	Total Líquido por Tipo de Peça.....	143
9.2.	Gráficos das previsões de séries temporais	144
9.2.1.	Total Líquido Global	144
9.2.2.	Total Líquido por Cliente – <i>Cluster 1</i>	149
9.2.3.	Total Líquido por Cliente – <i>Cluster 2</i>	154

9.2.4. Total Líquido por Tipo de Lavagem – <i>Cluster</i> 1	159
9.2.5. Total Líquido por Tipo de Lavagem – <i>Cluster</i> 2	164
9.2.6. Total Líquido por Tipo de Lavagem – <i>Cluster</i> 3	169
9.2.7. Total Líquido por Tipo de Lavagem – <i>Cluster</i> 4	174
9.2.8. Total Líquido por Tipo de Lavagem – <i>Cluster</i> 5	179
9.2.9. Total Líquido por Tipo de Lavagem – <i>Cluster</i> 6	184
9.2.10. Total Líquido por Tipo de Lavagem – <i>Cluster</i> 7	189
9.2.11. Total Líquido por Tipo de Lavagem – <i>Cluster</i> 8	194
9.2.12. Total Líquido por Tipo de Peça – <i>Cluster</i> 1	199
9.2.13. Total Líquido por Tipo de Peça – <i>Cluster</i> 2	204
9.3. Significância de coeficientes e modelos de previsão regressivos	209
9.3.1. Total líquido global	209
9.3.2. Total líquido por cliente – <i>cluster</i> 1	209
9.3.3. Total líquido por cliente – <i>cluster</i> 2	210
9.3.4. Total líquido por tipo de lavagem – <i>cluster</i> 1	210
9.3.5. Total líquido por tipo de lavagem – <i>cluster</i> 2	211
9.3.6. Total líquido por tipo de lavagem – <i>cluster</i> 3	211
9.3.7. Total líquido por tipo de lavagem – <i>cluster</i> 4	212
9.3.8. Total líquido por tipo de lavagem – <i>cluster</i> 5	212
9.3.9. Total líquido por tipo de lavagem – <i>cluster</i> 6	213
9.3.10. Total líquido por tipo de lavagem – <i>cluster</i> 7	213
9.3.11. Total líquido por tipo de lavagem – <i>cluster</i> 8	214
9.3.12. Total líquido por tipo de peça – <i>cluster</i> 1	214
9.3.13. Total líquido por tipo de peça – <i>cluster</i> 2	215
9.4. Significância de coeficientes e do modelo ARIMA	216
9.4.1. Total líquido global	216
9.4.2. Total líquido por cliente – <i>cluster</i> 1	216
9.4.3. Total líquido por cliente – <i>cluster</i> 2	216
9.4.4. Total líquido por tipo de lavagem – <i>Cluster</i> 1	217

9.4.5. Total líquido por tipo de lavagem – <i>Cluster 2</i>	217
9.4.6. Total líquido por tipo de lavagem – <i>Cluster 3</i>	217
9.4.7. Total líquido por tipo de lavagem – <i>Cluster 4</i>	217
9.4.8. Total líquido por tipo de lavagem – <i>Cluster 5</i>	218
9.4.9. Total líquido por tipo de lavagem – <i>Cluster 6</i>	218
9.4.10. Total líquido por tipo de lavagem – <i>Cluster 7</i>	218
9.4.11. Total líquido por tipo de lavagem – <i>Cluster 8</i>	218
9.4.12. Total líquido por tipo de peça – <i>Cluster 1</i>	219
9.4.13. Total líquido por tipo de peça – <i>Cluster 2</i>	219
9.5. Resíduos dos modelos de previsão.....	220
9.5.1. Total líquido global	220
9.5.2. Total líquido por cliente – <i>cluster 1</i>	221
9.5.3. Total líquido por cliente – <i>cluster 2</i>	222
9.5.4. Total líquido por tipo de lavagem – <i>cluster 1</i>	223
9.5.5. Total líquido por tipo de lavagem – <i>cluster 2</i>	224
9.5.6. Total líquido por tipo de lavagem – <i>cluster 3</i>	225
9.5.7. Total líquido por tipo de lavagem – <i>cluster 4</i>	226
9.5.8. Total líquido por tipo de lavagem – <i>cluster 5</i>	227
9.5.9. Total líquido por tipo de lavagem – <i>cluster 6</i>	228
9.5.10. Total líquido por tipo de lavagem – <i>cluster 7</i>	229
9.5.11. Total líquido por tipo de lavagem – <i>cluster 8</i>	230
9.5.12. Total líquido por tipo de peça – <i>cluster 1</i>	231
9.5.13. Total líquido por tipo de peça – <i>cluster 2</i>	232

Índice de Figuras

Figura 1 - Passos para a transformação numa organização orientada por dados.....	16
Figura 2 – Evolução do volume de dados digitais gerados entre 2010 e 2025 (zettabyte)	17
Figura 3 - Técnicas de Data Mining	23
Figura 4 - Exemplo de clustering, representando similaridades intra e inter clusters	26
Figura 5 - Procedimentos de análise de clusters.....	26
Figura 6 - Taxonomia dos algoritmos de clustering	27
Figura 7 - Utilização do algoritmo k-médias para formação de três clusters cujos pontos são identificados por triângulos, círculos e quadrados	29
Figura 8 - Regras de alinhamento para o algoritmo Euclidiano e Dynamic Time Warping.....	31
Figura 9 - Distorção (warping) entre duas séries temporais	32
Figura 10 - Representação do processo de cálculo do coeficiente silhueta $s(x_i)$ para um determinado objeto x_i de um cluster.....	34
Figura 11 - Exemplo de uma série temporal de padrão horizontal.....	37
Figura 12 - Exemplo de uma série temporal com padrão de tendência (crescente).....	38
Figura 13 - Exemplo de uma série temporal com padrão sazonal.....	38
Figura 14 - Exemplo de uma série temporal com padrão com tendência e sazonalidade	39
Figura 15 - Exemplo de uma série temporal com padrão cíclico	39
Figura 16 - Exemplo dos três métodos de previsão assinalados	41
Figura 17 - Imagens da lavandaria industrial Precioso & Moraes	47
Figura 18 - Variação do total líquido para os cinco clientes mais importantes em termos de faturação	57
Figura 19 - Aplicação dos algoritmos de clustering aos dados não estandardizados e estandardizados	58
Figura 20 - Variação do total líquido anual.....	64
Figura 21 - Distribuição do total líquido por cliente (correspondentes a 90 % do total líquido)	64
Figura 22 - Variação do total líquido por cliente (correspondente a 54,9 % do total líquido).....	65
Figura 23 - Distribuição do total líquido por tipo de lavagem (correspondente a 92,6 % do total líquido)	66
Figura 24 - Variação do total líquido por tipo de lavagem (correspondente a 92,6 % do total líquido) ..	66
Figura 25 - Distribuição do total líquido por tipo de peça (correspondente a 91,8 % do total líquido) ..	67
Figura 26 - Variação do total líquido por tipo de peça (correspondente a 79,3% do total líquido).....	68
Figura 27 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias sem estandardização, total líquido por cliente.	69
Figura 28 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias sem estandardização, total líquido por tipo de lavagem.....	69
Figura 29 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias sem estandardização, total líquido por tipo de peça.	70

Figura 30 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias com estandardização, total líquido por cliente.	70
Figura 31 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias com estandardização, total líquido por tipo de lavagem.	71
Figura 32 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias com estandardização, total líquido por tipo de peça.	71
Figura 33 - Variação do coeficiente silhueta na aplicação do algoritmo k-medóides sem estandardização, total líquido por cliente.	72
Figura 34 - Variação do coeficiente silhueta na aplicação do algoritmo k-medóides sem estandardização, total líquido por tipo de lavagem.	72
Figura 35 - Variação do coeficiente silhueta na aplicação do algoritmo k-medóides sem estandardização, total líquido por tipo de peça.	73
Figura 36 - Variação do coeficiente silhueta na aplicação do algoritmo k-medóides com estandardização, total líquido por cliente.	73
Figura 37 - Variação do coeficiente silhueta na aplicação do algoritmo k-medóides com estandardização, total líquido por tipo de lavagem.	74
Figura 38 - Variação do coeficiente silhueta na aplicação do algoritmo k-medóides com estandardização, total líquido por tipo de peça.	74
Figura 39 - Variação do total líquido por cliente. Cluster 1 a vermelho, cluster 2 a azul.	78
Figura 40 - Centróides dos clusters1 (vermelho) e 2 (azul).	78
Figura 41 - Variação do total líquido por cliente do cluster 2	79
Figura 42 - Distribuição dos tipos de peça dos clientes do cluster 2.....	80
Figura 43 - Distribuição dos tipos de lavagem dos clientes do cluster 2.	81
Figura 44 - Variação do total líquido por tipo de lavagem no cluster 1.	82
Figura 45 - Variação do total líquido por tipo de lavagem nos clusters2 a 8.....	83
Figura 46 - Distribuição do total líquido por tipo de lavagem	84
Figura 47 - Variação do total líquido por tipo de lavagem dos clusters1 a 8.....	84
Figura 48 - Distribuição do tipo de lavagens integrantes do cluster 1.	85
Figura 49 - Distribuição dos tipos de peça por tipos de lavagens integrantes do cluster 1.	86
Figura 50 - Distribuição dos tipos de peça por tipos de lavagens integrantes dos clusters2 a 8.....	87
Figura 51 - Variação do total líquido por tipo de peça.	89
Figura 52 - Variação do total líquido por tipo de peça.	89
Figura 53 - Variação do total líquido por tipo de peça.	90
Figura 54 - Distribuição dos tipos de lavagem pelos tipos de peça do cluster 1.....	91
Figura 55 - Distribuição dos tipos de lavagem dentro do tipo de peça do cluster 2.	92
Figura 56 - Distribuição por clientes do tipo de peça do cluster 2 (T-SHIRT).	93
Figura 57 - Representação gráfica da variação do total líquido no período em análise.....	94
Figura 58 - Representação gráfica da aplicação do método média à estimativa do total líquido	95
Figura 59 -Representação gráfica da aplicação do método ingénuo à estimativa do total líquido	96

Figura 60 - Representação gráfica da aplicação do método ingénuo sazonal à estimativa do total líquido	96
Figura 61 - Representação gráfica da aplicação do método deriva à estimativa do total líquido	97
Figura 62 - Representação gráfica da aplicação do modelo 1 à estimativa do total líquido	97
Figura 63 - Representação gráfica da aplicação do modelo 2 à estimativa do total líquido	98
Figura 64 - Representação gráfica da aplicação do modelo 3 à estimativa do total líquido	98
Figura 65 - Representação gráfica da aplicação do modelo 4 à estimativa do total líquido	99
Figura 66 - Representação gráfica da aplicação do método ARIMA à estimativa do total líquido	99

Índice de Tabelas

Tabela 1 - Vantagens e desvantagens dos métodos de clustering (Herman et al., 2022).....	30
Tabela 2 - Tabela de dados após processamento	54
Tabela 3 - Tipos de lavagem.....	54
Tabela 4 - Tipos de peça	55
Tabela 5 - Códigos de cliente	55
Tabela 6 - Total líquido mensal por cliente (95 clientes x 72 meses).....	56
Tabela 7 - Total líquido mensal por tipo de peça (43 peças x 72 meses)	56
Tabela 8 - Total líquido mensal por tipo de lavagem (20 lavagens x 72 meses).....	56
Tabela 9 – Coeficiente silhueta para diferentes valores do número k de clusters de clientes, tipos de lavagem e tipos de peça.	70
Tabela 10 - Coeficiente silhueta para diferentes valores do número k de clusters de clientes, tipos de lavagem e tipos de peça.	71
Tabela 11 - Coeficiente silhueta para diferentes valores do número k de clusters de clientes, tipos de lavagem e tipos de peça.	73
Tabela 12 - Coeficiente silhueta para diferentes valores do número k de clusters de clientes, tipos de lavagem e tipos de peça.	74
Tabela 13 - Quadro resumo dos resultados obtidos no clustering.	75
Tabela 14 - Quadro resumo do clustering dos clientes por total líquido.	75
Tabela 15 - Quadro resumo do clustering dos tipos de lavagem por total líquido.....	76
Tabela 16 - Quadro resumo do clustering dos tipos de peça por total líquido.	76
Tabela 17 - Distribuição dos clientes por cluster. A vermelho o cluster 1 e a azul o cluster 2.	77
Tabela 18 - Total líquido dos clientes do cluster 2 entre 2018 e 2022.	79
Tabela 19 – Distribuição dos tipos de lavagem por cluster	82
Tabela 20 - Distribuição dos tipos de peça por cluster. A vermelho o cluster 1 e a azul o cluster 2.	88
Tabela 21 - Resumo dos erros de previsão de cada método/modelo para o total líquido global.....	100
Tabela 22 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 1 do total líquido por cliente.....	101
Tabela 23 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 2 do total líquido por cliente.....	102
Tabela 24 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 1 do total líquido por tipo de lavagem	103
Tabela 25 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 2 do total líquido por tipo de lavagem	104
Tabela 26 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 3 do total líquido por tipo de lavagem	105
Tabela 27 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 4 do total líquido por tipo de lavagem	106
Tabela 28 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 5 do total líquido por tipo de lavagem	107

Tabela 29 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 6 do total líquido por tipo de lavagem	107
Tabela 30 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 7 do total líquido por tipo de lavagem	108
Tabela 31 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 8 do total líquido por tipo de lavagem	109
Tabela 32 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 1 do total líquido por tipo de peça	110
Tabela 33 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 2 do total líquido por tipo de peça	111
Tabela 34 - Análise global dos resultados dos modelos de previsão aplicados ao total líquido global e a cada cluster	112

1. Introdução

As empresas orientadas por dados são organizações que assentam as suas decisões estratégicas e operacionais na análise de dados quantitativos e qualitativos. Isto permite que as decisões sejam fundamentadas em informação concreta, ao contrário de intuições, reduzindo a subjetividade e aumentando a precisão e eficácia nas respostas aos desafios do mercado. Para se tornar uma empresa orientada por dados, uma organização precisa de uma infraestrutura tecnológica capaz, com sistemas que permitam a recolha, o processamento e a análise de dados. Este investimento permitirá transformar grandes volumes de dados em informações e conhecimento, o que se pode traduzir em vantagens competitivas. O uso eficiente de dados permite a uma organização aprofundar o conhecimento sobre o comportamento dos clientes, otimizar operações internas e antecipar tendências de mercado (Carl. Anderson, 2015; Assur & Rowshankish, 2022; Hupperz et al., 2021).

Para além do referido no parágrafo anterior, para o sucesso desta abordagem é essencial a criação de uma cultura organizacional que valorize o uso de dados. As empresas orientadas por dados promovem a transparência e a acessibilidade dos dados a toda a organização, de forma a facilitar a colaboração interna e assegurar que todas as áreas de negócio podem tomar decisões informadas. Este modelo de funcionamento promove um alinhamento estratégico robusto e potencia a capacidade de inovação, ao permitir que cada departamento identifique oportunidades e problemáticas de forma célere e fundamentada. Para além disto, estas organizações conseguem monitorizar o impacto das suas ações em tempo real, ajustando rapidamente as suas estratégias com base em métricas de desempenho objetivas. No cenário económico atual, caracterizado pela rápida transformação digital, as empresas que adotam uma orientação por dados estão melhor posicionadas para se adaptarem e prosperarem, utilizando os dados não apenas para conhecer o seu desempenho operacional, prever receitas, tendências e comportamentos futuros, mas também para a definição de estratégias que permitam o crescimento do modelo de negócio (Carl. Anderson, 2015; Tan et al., 2019)

A transformação de uma empresa numa organização orientada por dados, com a consequente digitalização de processos produtivos e operacionais, produz volumes significativos de dados, criando novas oportunidades para otimização e tomada de decisões mais informadas. Neste contexto, o *data mining* destaca-se como uma ferramenta para extrair padrões, prever tendências e identificar anomalias, especialmente em ambientes industriais (Carl. Anderson, 2015).

A aplicação de *data mining* oferece um novo horizonte para otimização de processos, gestão de *stock*, previsão de procura e manutenção preventiva. Através da análise detalhada de grandes volumes de dados, gerados diariamente, é possível identificar padrões que contribuem diretamente para a melhoria da eficiência e para o conhecimento da operação da

empresa. Além disso, o uso de algoritmos preditivos pode auxiliar na antecipação de falhas em processos operacionais, na alocação de recursos e no melhoramento do serviço ao cliente (Gladju et al., 2022; Regin & Suman Rajest, 2023; Tan et al., 2019).

O *data mining* recorre, entre outras, à ferramenta de análise de dados designada por *clustering*, que é uma técnica de análise de dados que agrupa elementos com características semelhantes em conjuntos ou *clusters*, facilitando a identificação de padrões e a extração de informação a partir de grandes volumes de dados. Em empresas, o *clustering* é utilizado para segmentar clientes, identificar padrões de comportamento de compra, otimizar a logística e personalizar a oferta de produtos e serviços. Através da aplicação destes algoritmos uma empresa consegue identificar conjuntos de clientes ou operações com perfis e necessidades específicas, o que melhora o processo de tomada de decisão, apoiada em informação mais detalhada e consistente (Huang et al., 2007; Li et al., 2022).

Ao implementar técnicas de *clustering* uma empresa compreende melhor as suas áreas de negócio, podendo também antecipar as necessidades e preferências dos seus clientes. Esta técnica de análise permite dar prioridade a determinadas ações e desenvolver estratégias específicas para cada segmento identificado, seja na criação de campanhas de *marketing* mais direcionadas, na adaptação do serviço ao cliente ou na otimização da gestão de *stock*. Para além da informação obtida sobre o desempenho da empresa, podem ainda ser extrapolados para o futuro, através de ferramentas de previsão, os conhecimentos obtidos através da aplicação do *clustering* (Garibaldi et al., 2006; Huang et al., 2007; Li et al., 2022).

A previsão de desempenho futuro, através de modelos de previsão de séries temporais, permite às empresas projetar tendências e padrões futuros com base em dados históricos. Em termos de operação, esta abordagem é importante para prever a procura, para a gestão de *stock* e para a antecipação das necessidades de recursos, permitindo uma alocação mais precisa e eficiente. Modelos de previsão de séries temporais, desde os mais simples, a métodos regressivos e auto-regressivos, analisam padrões sazonais, ciclos e variações, permitindo a elaboração de previsões que permitem às empresas o ajuste das operações, de acordo com a procura esperada. Com estas previsões, a empresa pode minimizar o desperdício, melhorar a eficiência operacional e reduzir custos, ajustando proactivamente a sua capacidade de resposta (Montgomery et al., 2015; Mondal et al., 2014).

Particularizando na gestão da receita de uma empresa, a previsão baseada em séries temporais é igualmente importante, pois permite antecipar flutuações e planejar investimentos futuros com base em projeções baseadas em dados históricos. A análise de séries temporais de faturação permite identificar padrões de vendas, períodos de pico e sazonalidade, bem como prever o impacto de fatores externos, como alterações no mercado ou no comportamento dos consumidores. Com previsões de receita mais precisas, as empresas

podem gerir o fluxo de caixa com maior segurança, identificar oportunidades de crescimento e ajustar as suas estratégias de *marketing* e vendas para otimizar o desempenho financeiro (Huang et al., 2007; Regin & Suman Rajest, 2023).

1.1. Aplicação de *data mining* à análise da atividade de uma lavandaria industrial

Uma lavandaria industrial é uma empresa do setor têxtil que oferece serviços especializados de lavagem e acabamento de produtos têxteis, integrando-se no processo produtivo industrial. Estes serviços englobam, por exemplo, lavagem, secagem, tratamento, dobragem e transporte de artigos têxteis. O setor responsável pelo tratamento em larga escala de roupas e têxteis, é uma área que, tradicionalmente, possui uma forte dependência de processos operacionais contínuos e eficientes. Contudo, apesar do crescente uso de tecnologias de registo de dados, muitas empresas ainda carecem de mecanismos sofisticados de análise para maximizar o uso eficiente dos seus recursos e minimizar custos operacionais.

Esta dissertação tem como objetivo explorar a aplicação de técnicas de *data mining* no contexto de uma lavandaria industrial, contribuindo para o processo de transformação da mesma numa empresa orientada por dados, para o conhecimento do desempenho da lavandaria e para o desenvolvimento de modelos de previsão de desempenho futuro.

Um estudo sobre a aplicação de técnicas de *data mining* à análise da atividade desta lavandaria industrial é relevante por permitir a extração de padrões e conhecimentos úteis a partir de grandes volumes de dados operacionais. Este estudo permite obter informações importantes para a gestão da empresa, ao analisar os dados que são diariamente recolhidos. É de notar que, antes deste estudo, não era feita nenhuma análise aos dados registados. O trabalho de análise de dados contribui para o apoio à decisão, bem como para o planeamento estratégico da empresa, prevendo resultados, antecipando desafios e promovendo um crescimento sustentado do modelo de negócio.

A previsão de dados correspondentes a séries temporais pode contribuir para o sucesso da lavandaria industrial em consideração. Ao analisar e antecipar a procura futura com base nos dados históricos, a empresa pode otimizar as suas operações, gerir os recursos de forma mais eficiente e preparar-se para enfrentar futuras necessidades. A aplicação de metodologias de previsão permite transformar grandes volumes de dados em conhecimentos estratégicos, promovendo o crescimento sustentável e assegurando a competitividade da empresa no setor. A previsão de séries temporais pode ser importante para o funcionamento da lavandaria industrial por diversas razões:

- **Planeamento de Recursos:** A previsão de séries temporais permite à lavandaria antecipar a procura futura pelos seus serviços, possibilitando uma gestão mais eficaz dos recursos, tais como equipamentos, produtos químicos, mão de obra e logística. Por exemplo, ao prever aumentos sazonais no volume de peças a tratar, a organização pode garantir que dispõe de pessoal e equipamento suficientes para dar resposta à procura, evitando, assim, estrangulamentos operacionais.
- **Gestão de *stock*:** Com base na previsão dos volumes de trabalho a partir dos padrões históricos, a lavandaria pode ajustar o seu *stock* de produtos, como detergentes, amaciadores e outros produtos de limpeza especializados. Isto permite não só reduzir os custos associados ao excesso de inventário, mas também evitar a falta de materiais críticos para o processo produtivo.
- **Otimização de Processos:** A previsão possibilita a identificação de padrões sazonais ou tendências na procura de diferentes tipos de lavagens e peças. Com estas informações, a lavandaria pode otimizar a programação dos serviços, ajustando os processos operacionais de forma a maximizar a eficiência durante os períodos de maior procura e a minimizar o desperdício nos períodos de menor atividade.
- **Definição de Metas e Estratégias:** A previsão é igualmente crucial para a definição de metas realistas de crescimento e rentabilidade. Com base nos dados históricos, a lavandaria pode projetar a sua receita futura e ajustar as suas estratégias comerciais e de *marketing*, como a implementação de campanhas direcionadas para clientes específicos ou serviços com maior procura, de forma a maximizar o seu desempenho financeiro.
- **Prevenção de Problemas Operacionais:** As séries temporais podem também ser utilizadas para identificar anomalias ou padrões inesperados, que possam indiciar potenciais problemas, tais como uma diminuição súbita da procura ou ineficiências nos processos operacionais. Ao detetar estas anomalias atempadamente, a lavandaria pode agir de forma proativa, mitigando os problemas antes que estes comprometam a continuidade ou a qualidade do serviço prestado.
- **Competitividade:** A vantagem competitiva de uma organização está diretamente relacionada com a sua capacidade de tomar decisões informadas com base em dados fiáveis. Uma lavandaria industrial que utiliza previsões de séries temporais estará mais apta a antecipar mudanças no mercado, ajustando rapidamente a sua oferta e mantendo-se competitiva ao prestar serviços de forma mais eficiente, com menores custos operacionais e maior capacidade de resposta.
- **Previsão de receita:** A previsão de receita futura numa lavandaria industrial pode ser um instrumento essencial para a gestão financeira e estratégica da empresa.

Baseando-se na análise de séries temporais dos dados históricos de faturação, como o volume de serviços prestados a cada cliente, tipos de lavagens e peças, é possível estimar com maior precisão os fluxos de receita nos períodos subsequentes. Esta capacidade preditiva permite à organização alinhar melhor os seus investimentos, otimizar a alocação de recursos e ajustar as suas estratégias de preços e *marketing*, garantindo uma maior estabilidade financeira e promovendo o crescimento sustentado do negócio. A previsão de receita, ao mitigar a incerteza associada à variabilidade da procura, capacita a empresa a tomar decisões informadas, reduzindo os riscos e maximizando a rentabilidade.

Dado o prazo para realização desta dissertação, a análise foi deliberadamente circunscrita à previsão da receita futura. Embora outros aspetos operacionais da lavandaria industrial, como a alocação de recursos, a gestão de *stock* ou a análise de padrões de consumo, possam também ser objeto de previsão e otimização, o foco desta investigação recaiu exclusivamente sobre a estimativa dos fluxos de receita. Esta escolha visa assegurar uma abordagem concentrada e aprofundada, permitindo explorar com maior rigor as metodologias de previsão de séries temporais aplicadas à receita, sem dispersar a análise por múltiplas variáveis que, embora relevantes, não poderiam ser adequadamente tratadas dentro das limitações temporais e de dados disponíveis para o estudo.

Assim, a metodologia utilizada neste estudo assenta na aplicação de técnicas de *clustering* no sentido de agrupar os dados recolhidos pela lavandaria industrial em grupos com similaridades, realizando uma análise a cada um desses grupos e numa segunda fase, no desenvolvimento de modelos de previsão da receita futura da empresa, com base em modelos de previsão de séries temporais.

O tema desta dissertação surge da relação familiar do autor com os proprietários da lavandaria industrial alvo de análise. O conhecimento sobre a operação da lavandaria industrial, permitiu identificar oportunidades de melhoria ao nível das informações sobre os clientes, dos comportamentos destes, do volume e tipologia de serviços, da faturação e de muitas outras características do desempenho operacional desta empresa. A gerência da empresa tem, obviamente, conhecimentos sólidos sobre o desempenho e a operação da lavandaria, caso contrário, não teriam um modelo de negócio de sucesso, o que é inegável. No entanto, este conhecimento é empírico, assente em décadas de trabalho neste modelo de negócio. A motivação deste estudo é complementar esse conhecimento empírico com uma análise científica de dados, de forma a aperfeiçoar a operação e promover o crescimento da lavandaria industrial.

O impacto que este estudo tem a nível pessoal é, primariamente, a conclusão do Mestrado em Gestão. No entanto, o impacto vai muito para além do nível primário. A nível profissional, pode-se dividir em duas linhas de ação. Uma focada na atividade profissional do autor, que lida com dados diariamente, em que o desenvolvimento de conhecimentos na área da análise de dados, permitirá a melhoria no âmbito das suas funções, contribuindo de forma significativa para o planeamento e gestão das atividades de sua responsabilidade. A segunda linha, funde-se com o impacto a nível pessoal do autor, pois o desempenho profissional da gerência da lavandaria é também o desempenho da família do autor, o que faz com que qualquer ganho de conhecimento, seja uma melhoria sempre em duas vertentes.

1.2. Organização da dissertação

Esta dissertação está organizada nos seguintes capítulos:

O Capítulo 2 faz uma revisão de literatura existente sobre os diferentes temas no âmbito desta dissertação.

O capítulo 3 apresenta os métodos de *clustering* e de previsão considerados neste trabalho.

O Capítulo 4 apresenta a lavandaria industrial Precioso & Moraes, Lda., foco sobre o qual incide esta dissertação de mestrado, abordando a questão de a empresa não ser orientada por dados.

O Capítulo 5 descreve a metodologia aplicada neste estudo, nomeadamente na recolha e processamento dos dados, na aplicação da técnica de *clustering* e na previsão de séries temporais aos dados recolhidos à data de início deste trabalho.

O Capítulo 6 apresenta os resultados obtidos pela aplicação das ferramentas de análise de dados e interpreta os mesmos. Inicialmente é feita uma análise primária aos dados recolhidos, seguindo-se uma análise aos resultados do *clustering*, caracterizando todos os *clusters* obtidos e por fim analisado o resultado da aplicação dos modelos de previsão de séries temporais.

O Capítulo 7 faz um breve resumo dos capítulos anteriores, apresentando as conclusões obtidas neste estudo, bem como, uma abordagem a possíveis futuros trabalhos

de continuação desta análise e do processo de transformação desta lavandaria, numa empresa orientada por dados.

O estilo seguido na organização e à citação das referências bibliográficas é o definido pela American Psychological Association (APA) de acordo com (American Psychological Association, 2020).

2. Revisão de literatura

Um dos campos em que o *data mining* é comumente utilizado é na área das previsões económicas e no *marketing*, existindo vários estudos sobre a aplicação de técnicas de *clustering* e de previsão. Na área da modelação macroeconómica Garibaldi et al. (2006) descreve variáveis macroscópicas com base no comportamento de uma grande coleção de entidades microeconómicas. Khan et al. (2020a) explora como o *data mining* e o *machine learning* ajudam a prever procura futura com base em dados de vendas recolhidos de diversas fontes. O estudo descreve o processo de previsão de procura, desde a recolha de dados brutos até à determinação de procuras semanais, mensais e trimestrais. Um artigo publicado por Tsiu et al. (2025) refere que investigação recente mostra que tecnologias como a *data mining* e a inteligência empresarial têm grande impacto nas PME, melhorando a gestão de clientes, cadeias de abastecimento e previsão financeira, bem como o planeamento estratégico e a inovação. Dado o papel crucial das PME na economia global, considera-se fundamental explorar a utilização destas ferramentas para fortalecer a sua competitividade. Em termos de literatura existente sobre o uso da inteligência empresarial nas PME, é referido que a mesma é limitada, pois a maioria dos estudos foca-se em grandes empresas. Um estudo por Makridakis et al. (2009) aborda a importância das previsões para decisões económicas e empresariais, propondo uma nova abordagem para lidar com o futuro, apresentando um quadro que ajuda a enfrentar a incerteza e os limites das previsões em decisões orientadas para o futuro. Kolkova (2020) aborda a importância da previsão precisa de variáveis de negócios como um elemento essencial para a competitividade das empresas, especialmente num ambiente globalizado e digitalizado. O objetivo do estudo é avaliar a precisão de modelos que preveem receitas no setor de serviços, identificando se o uso de determinados critérios pode levar à adoção de modelos específicos para melhorar previsões competitivas. O trabalho de Kumar (2020) trata da aplicação de *data mining* como uma ferramenta amplamente utilizada na engenharia, ciências, finanças e *marketing*, para ajudar a resolver desafios específicos dessas áreas, referindo que um sistema de suporte à decisão baseado em *data mining* melhora o desempenho organizacional ao analisar a realidade do mercado. A segmentação de clientes, através da utilização de algoritmos de *clustering*, é uma área em que Huang et al. (2007) desenvolve um trabalho, referindo que esta é amplamente utilizada para direccionar um mercado menor, sendo útil para os decisores alcançarem todos os clientes de forma eficaz com uma única estratégia de *marketing*. Já Alves Gomes & Meisen (2023) aborda a crescente importância do *marketing* orientado para o cliente nas últimas décadas, especialmente com o advento das estratégias individualizadas de *marketing*, como as aplicadas no *e-commerce*, que tornam obsoletas as tradicionais abordagens de *marketing* em massa, referindo que essa estratégia exige uma compreensão profunda dos interesses e

motivações de cada cliente, sendo a segmentação um método comumente utilizado para esse efeito, que tem evoluído de forma constante.

No entanto, existem outras áreas de atividade, onde ferramentas de análise de dados são utilizadas para melhoria dos serviços e das diferentes atividades. No campo da energia renovável, Sweeney et al. (2020) aborda a crescente importância da previsão de energia renovável gerada por vento e solar, à medida que sua participação na matriz energética aumenta, oferecendo uma visão geral do estado da arte em previsões para essas fontes, cobrindo abordagens de modelagem estatística e física para diferentes escalas de tempo, desde minutos até dias, tanto em previsões determinísticas quanto probabilísticas. Li et al. (2022) apresenta um modelo inovador para previsão de geração de energia fotovoltaica, com foco em melhorar a precisão no planeamento económico e confiável da produção de energia. O modelo propõe uma previsão multi-etapas baseada numa combinação de redes generativas adversarias de séries temporais, algoritmos de *clustering* K-medóides com distorção dinâmica suave de tempo (soft-DTW) e uma rede neuronal híbrida. No campo da aquacultura, Gladju et al. (2022) aborda como os setores de aquicultura e pesca estão a utilizar de forma inovadora a tecnologia da informação, ciência de dados e inteligência artificial para atender à crescente procura humana por peixe e frutos do mar ricos em nutrientes. Estas tecnologias estão a ser progressivamente integradas nas práticas de pesca e cultivo de peixes para intensificar a produção, explorar de forma sustentável os recursos pesqueiros naturais e mecanizar atividades relacionadas. Sistemas de *data mining* e *machine learning* estão a ser desenvolvidos para processar grandes volumes de dados complexos, realizando tarefas inteligentes como análise de relações causa-efeito, previsão de problemas e fornecimento de soluções precisas para o cultivo e captura de peixes. No campo da medicina, Chen et al. (2023) aborda o uso crescente da tecnologia da informação na saúde, o que resulta numa extensa recolha de dados, que podem ser utilizados para melhorar o atendimento ao paciente e a gestão de doenças crónicas, resumindo as investigações anteriores que utilizaram métodos de *data mining* no contexto de doenças crónicas.

O resto deste capítulo está organizado em várias secções que reveem e relevam individualmente a temática das empresas orientadas por dados e o *data-mining*.

2.1. Empresa orientada por dados

Os dados desempenham um papel crucial no desenvolvimento económico, servindo de base para a criação de novos produtos e serviços, além de impulsionarem a produtividade e a eficiência em todos os setores. Através da utilização inteligente de dados, é possível personalizar produtos e serviços, melhorar a formulação de políticas públicas e modernizar os serviços estatais. Para empresas em fase de arranque e Pequenas e Médias Empresas, ou

seja, não apenas para grandes empresas, os dados são um recurso valioso, permitindo-lhes inovar e competir. Além disso, a disponibilidade de dados é fundamental para o treino de sistemas de inteligência artificial, que evoluem do simples reconhecimento de padrões para técnicas de previsão avançadas, apoiando decisões mais eficazes (Comissão Europeia, 2020).

As organizações precisam de integrar tecnologias de análise, como a análise preditiva e prescritiva, para explorar o potencial dos dados e obter informações e descobrir oportunidades de negócio. O termo "orientado por dados" é frequentemente utilizado para destacar o papel central dos dados na tomada de decisões e na estratégia. Essa abordagem permite que as empresas tomem decisões mais informadas, identificando tendências, otimizando operações e explorando novas oportunidades de crescimento (Hupperz et al., 2021).

Muitas organizações acreditam que, ao gerar inúmeros relatórios ou ter vários *dashboards*, são orientadas por dados. Embora essas atividades façam parte da operação, elas tendem a ser retrospectivas, ou seja, apenas apresentam fatos passados ou presentes, geralmente sem fornecer o contexto adequado, sem explicações sobre as causas dos eventos e sem recomendações sobre os próximos passos. Em resumo, indicam o que aconteceu, mas não são prescritivas. Assim, o seu valor é limitado (Carl. Anderson, 2015).

A gestão de mudanças orientada por dados torna os processos de mudança mais mensuráveis e bem-sucedidos. Isso ocorre, em grande parte, devido à maior transparência, a mais oportunidades de participação e a um maior senso de controlo entre os envolvidos. A nova gestão de mudanças orientada por dados integra uma camada adicional de informação. Os dados fornecem argumentos convincentes que os executivos utilizam para demonstrar a necessidade da mudança e justificar os meios escolhidos. Em organizações com maior domínio de dados, a probabilidade de obter apoio generalizado para a mudança desde o início é significativamente maior (Crummenerl et al., 2023).

Segundo Assur & Rowshankish (2022) uma empresa orientada por dados, é definida por sete características:

Integração de dados

Os dados estão integrados em todas as decisões, interações e processos - As organizações conseguem tomar decisões mais informadas e automatizar atividades básicas do dia-a-dia, bem como decisões que ocorrem regularmente. Os principais facilitadores para este processo são uma visão e estratégia de dados, tecnologia facilitadora e a literacia de dados e cultura orientada por dados.

Processamento de dados

Os dados são processados e entregues em tempo real - Redes extensas de dispositivos conectados coletam e transmitem dados em tempo real. Tecnologias avançadas transformam significativamente a geração, processamento, análise e visualização dos dados, proporcionando a disponibilização de informação de forma mais rápida e importante. Os principais potenciadores deste processo são uma visão da arquitetura empresarial completa, dispositivos de *edge-computing*¹ mais poderosos, infraestruturas de conectividade avançadas e computação em memória.

Armazenamento de dados

Armazenamentos de dados flexíveis permitem dados integrados e prontos para uso – A utilização de uma variedade de tipos de bases de dados, incluindo bases de dados de séries temporais, bases de dados de grafos e bases de dados NoSQL², o que permite formas mais flexíveis de organizar os dados. Para potenciar este processo é necessária uma arquitetura de dados moderna e o desenvolvimento de modelos de dados e gémeos digitais para replicar os sistemas do mundo real.

Modelo operacional de dados

O modelo operacional de dados trata os dados como um produto - Os ativos de dados são organizados e suportados como produtos, independentemente de serem utilizados por equipas internas ou clientes externos. Os produtos de dados evoluem continuamente de maneira ágil para atender às necessidades dos consumidores, aproveitando operações de dados e processos e ferramentas de integração e entrega contínuas. Para facilitar este processo é necessária uma estratégia de dados, entender as fontes e tipos de dados e a existência de um proprietário e uma equipa de dados.

Chief data officer

O papel do *Chief Data Officer* é expandido para gerar valor - Os *Chief Data Officers* e as suas equipas operam como uma unidade de negócios com responsabilidades

¹ *Edge computing* é uma estrutura de computação distribuída que aproxima as aplicações empresariais das fontes de dados, como dispositivos *Internet of Things* (rede coletiva de dispositivos conectados e à tecnologia que facilita a comunicação entre os dispositivos e a nuvem, bem como entre os próprios dispositivos) ou servidores locais de *edge*.

² Estes tipos de bases de dados são otimizados especificamente para aplicações que requerem modelos de dados flexíveis, grandes volumes de dados e baixa latência, conseguindo isso ao relaxar algumas das restrições de consistência dos bancos de dados relacionais.

financeiras. Eles colaboram com as equipas de negócios para criar novas formas de utilizar dados, desenvolver e integrar uma estratégia de dados empresarial e gerar receitas por meio da monetização e partilha de dados. Os potenciadores deste processo são a literacia de dados, um modelo económico que reconheça e atribua dados e custos, desenvolvimento de talento e uma aposta na experimentação e inovação.

Ecossistema de dados

A adesão a ecossistemas de dados torna-se a norma - Organizações grandes e complexas utilizam plataformas de partilha de dados para colaborar em projetos orientados por dados, tanto internamente quanto externamente. Empresas orientadas por dados participam na economia de dados para combinar informações e gerar perceções importantes. Para tal é necessária a adoção de modelos de dados comuns e o desenvolvimento de alianças e acordos de partilha.

Gestão de dados

A gestão de dados é priorizada e automatizada para garantir privacidade, segurança e resiliência - As mentalidades organizacionais devem mudar completamente para tratar a privacidade, a ética e a segurança dos dados como áreas de competência necessária, impulsionadas pela evolução das expectativas regulatórias. Os principais potenciadores deste processo são a elevação da importância da segurança dos dados, o crescimento da consciencialização dos consumidores, a adoção de tecnologias automáticas de administração de bases de dados e a adoção de ferramentas de resiliência e armazenamento de dados baseadas na nuvem.

A transformação bem-sucedida de uma empresa depende do envolvimento ativo dos colaboradores, e os gestores de mudança dedicam-se a explicar e superar resistências internas, podendo ser dividida em quatro passos, conforme a figura seguinte, nomeadamente a definição de uma estratégia, a construção de confiança, a modernização e a ativação. Os dados são fundamentais nesse processo, esclarecendo o propósito da mudança, monitorizando o progresso e ajustando o rumo quando necessário. Com análises modernas, os dados tornam-se essenciais para ajudar gestores a convencer e engajar os colaboradores de forma eficaz (Crummenerl et al., 2023).

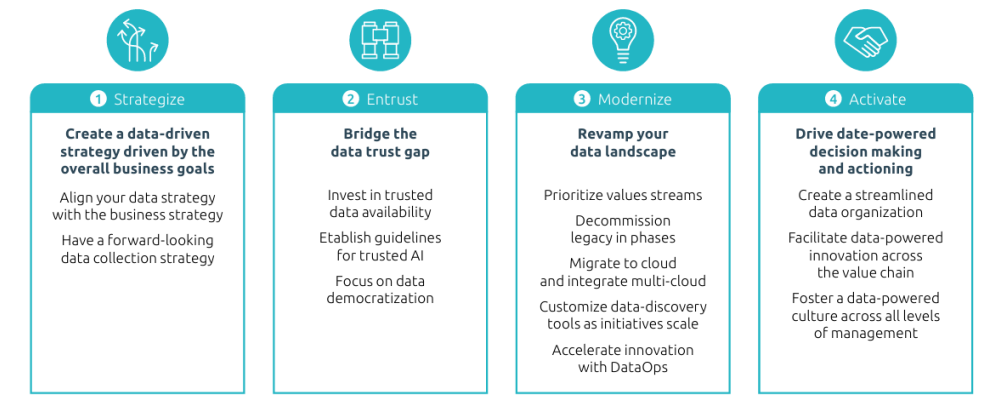


Figura 1 - Passos para a transformação numa organização orientada por dados (Crummenerl et al., 2023)

Fomentar a confiança nos dados envolve tanto a cultura corporativa quanto a tecnologia. O sucesso nas mudanças baseadas em dados fortalece a confiança dos colaboradores no seu uso. Para manter essa confiança, é essencial informar os colaboradores sobre os dados e ferramentas utilizadas e como eles contribuem para os objetivos do projeto. Criar um ambiente de dados avançado é fundamental para o êxito das mudanças. Pequenas adaptações técnicas, quando os analistas de dados colaboram com as tecnologias da informação, podem garantir que os sistemas atendam às necessidades analíticas. Ferramentas como *dashboards* de relatórios, especialmente com funções de inteligência artificial, ajudam a identificar e aplicar as medidas mais eficazes. Se a equipa de mudança cumprir as estratégias e construir um ambiente de confiança e modernidade nos dados, coloca o projeto no rumo certo. Com o tempo, a base de dados expande-se, e as decisões — também dos colaboradores — tornam-se mais sólidas, tornando o processo de mudança mais suave. Desta forma, os líderes de mudança tornam-se pioneiros de uma nova cultura corporativa orientada por dados (Crummenerl et al., 2023).

2.2. Data mining

De acordo com Jain (2010) os avanços na tecnologia de sensores e armazenamento, juntamente com o aumento das aplicações ligadas à pesquisa na internet, imagens digitais e videovigilância, resultam na criação de grandes volumes de dados complexos. A maior parte desses dados é digital, permitindo o desenvolvimento de técnicas automáticas para a sua análise. A diversidade dos dados (textos, imagens, vídeos) também cresceu, em consequência da maior disponibilidade de câmaras de vídeo digitais e etiquetas de identificação por rádio frequência (RFID), contribuindo desta forma para milhões de sensores que transmitem dados. E-mails, blogs e páginas da web geram *terabytes* de novos dados diariamente, muitos dos quais são não estruturados, aumentando assim a complexidade da sua análise.

O volume de dados digitais gerados anualmente tem vindo a crescer rapidamente, com previsões de crescimento exponencial contínuo nos próximos anos. A Figura 1 apresenta a evolução da produção de dados digitais entre 2010 e 2020, com as estimativas relativas a 2021 a 2025 (Statista, 2024).

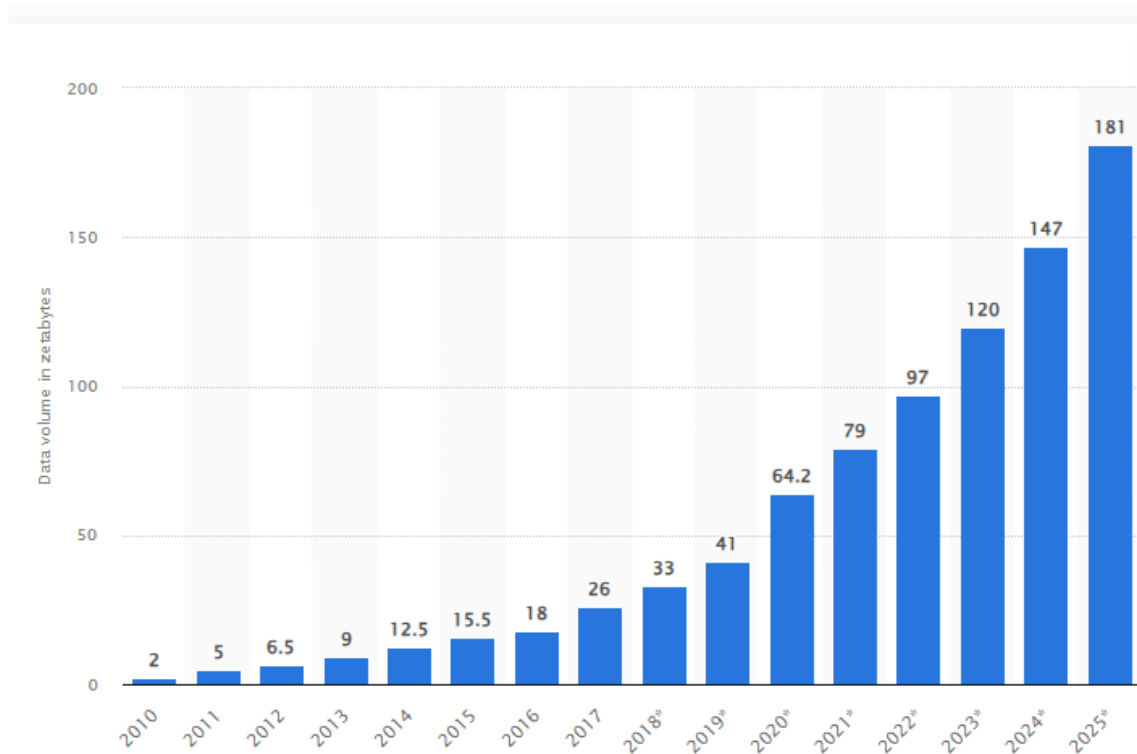


Figura 2 – Evolução do volume de dados digitais gerados entre 2010 e 2025 (zettabyte) (Statista, 2024)

A capacidade de extrair informações significativas de grandes volumes de dados é crucial para a tomada de decisões em diversas áreas da sociedade, como negócios, ciência, medicina, governo e para indivíduos. No entanto, o volume, a variedade e a velocidade desses dados são tão elevados que a análise manual se torna inviável. Por isso, há uma necessidade crescente de ferramentas automatizadas que possibilitem a extração de informações úteis de *big data*, apesar dos desafios relacionados à sua escala e diversidade (Tan et al., 2019). Para permitir lidar com esta problemática, foram criadas várias metodologias de organização e modelação de dados, motivadas por aplicações em diagnóstico, educação, previsão e outros campos. A definição, avaliação e aplicação dessas metodologias pertencem ao campo do *machine learning*, que se tornou uma subárea importante da ciência da computação e da estatística devido ao seu papel essencial no mundo moderno (Rodriguez et al., 2019).

O *machine learning* distingue-se entre dois tipos relevantes para a análise de dados: aprendizagem supervisionada e não supervisionada. A aprendizagem supervisionada requer dados categorizados e conhecimento prévio sobre os padrões a serem detetados, sendo comumente utilizado para tarefas de regressão e classificação. Em contraste, a aprendizagem

não supervisionada não necessita de dados categorizados e procura encontrar padrões desconhecidos nos dados, sendo frequentemente empregue no reconhecimento de padrões (Wegmann, Zipperling, et al., 2021).

Como a maioria dos dados encontra-se registada de forma arbitrária, o seu agrupamento pode ser difícil tendo em conta a falta de conhecimento prévio sobre as características dos objetos de dados. Sem este agrupamento ou classificação, os dados podem tornar-se instâncias de aprendizagem não supervisionada. A análise de agrupamento de dados, especialmente no campo do *data mining*, onde o objetivo é extrair informação útil de grandes volumes de dados, tem desempenhado ao longo do tempo um papel vital na organização e classificação adequada desses dados (Ezugwu et al., 2021).

O *data mining*, segundo (Tan et al., 2019), combina métodos tradicionais de análise de dados com algoritmos sofisticados para processar essa abundância de dados, sendo de destacar algumas das possíveis aplicações:

- **Negócios e Indústria** - A recolha de dados no ponto de venda (*scanners* de códigos de barras, identificação por radiofrequência - RFID, e tecnologia de cartões inteligentes) permite aos retalhistas recolher dados atualizados sobre as compras dos clientes nas caixas registadoras das lojas. Os retalhistas podem utilizar esta informação, juntamente com outros dados críticos para o negócio, como registos de servidores *web* de *websites* de *e-commerce* e registos de atendimento ao cliente dos *call centers*, para compreender melhor as necessidades dos clientes e tomar decisões empresariais mais informadas.

As técnicas de *data mining* podem apoiar uma ampla gama de aplicações de inteligência empresarial, como perfis de clientes, *marketing* direcionado, gestão de fluxo de trabalho, disposição das lojas, deteção de fraudes e automatização de compras e vendas. O *data mining* também pode ajudar os retalhistas a responder a perguntas cruciais para o negócio, como "Quem são os clientes mais lucrativos?", "Que produtos podem ser vendidos em conjunto ou adicionalmente?" e "Qual é a previsão de receita da empresa para o próximo ano?". Estas questões inspiraram o desenvolvimento de técnicas de *data mining*, como a análise de associações (Tan et al., 2019). Nos serviços financeiros o *data mining* e a análise preditiva são fundamentais, permitindo melhorar a deteção de fraudes e a avaliação de riscos. Estas técnicas analisam dados históricos e padrões para prever comportamentos suspeitos, ajudando a tomar decisões mais informadas e a prevenir riscos e fraudes, garantindo a conformidade regulatória e a proteção de ativos (Javaid, 2024).

As atividades empresariais de uma organização envolvem a evolução contínua dos comportamentos de grupos e os riscos financeiros que surgem são uma

consequência direta dessas mudanças constantes. A identificação de potenciais riscos financeiros pode ser vista como um desafio estratégico entre aqueles com acesso a informações confidenciais e aqueles que as procuram ativamente. As técnicas de mineração de grandes volumes de dados, conhecidas como *big data mining*, oferecem um potencial significativo para melhorar a estrutura organizacional, aumentar a eficiência da gestão e reduzir substancialmente o risco financeiro (Regin & Suman Rajest, 2023).

O campo da pesca e da aquacultura providencia meios de subsistência para 59,5 milhões de pessoas em todo o mundo. Com o crescimento contínuo e rápido deste setor, há agora um volume sem precedentes de dados disponíveis sobre pesca e aquacultura. Este aumento de dados exige a adoção de tecnologias inteligentes, baseadas em sistemas inovadores de dados e de análise, para melhorar a sustentabilidade económica, social e ambiental ao longo das cadeias de abastecimento de pescado. Atualmente, existe uma variedade de tecnologias de recolha e processamento de dados, como imagens de satélite de alta resolução, redes de sensores múltiplos in situ, sistemas de identificação automática, *big data*, aprendizagem automática, inteligência artificial e a Internet das Coisas, que permitem a monitorização total das pescas e a aquacultura de precisão. Por exemplo, os algoritmos de aprendizagem automática podem prever e quantificar o esforço de pesca com uma resolução espacial e temporal sem precedentes, ao monitorizar os movimentos dos navios de pesca, o uso de equipamentos e os *stocks* de peixe. Da mesma forma, os sensores e o processamento de dados podem ajudar a identificar padrões em sistemas de aquacultura, apresentando estratégias de gestão rentáveis para os produtores (Gladju et al., 2022).

- Setor têxtil – O *machine learning* está a transformar progressivamente a indústria têxtil ao permitir a tomada de decisões baseada em dados, a automatização de processos e o desenvolvimento de sistemas inteligentes. Os modelos de *machine learning* conseguem prever defeitos, otimizar parâmetros e personalizar produtos, resultando em benefícios significativos como melhor qualidade, redução de custos de produção e menor tempo de entrega. O *data mining*, por sua vez, é especialmente importante para identificar informações ocultas e melhorar a eficiência dos processos produtivos, lidando eficazmente com conjuntos de dados complexos. A integração de técnicas de *machine learning* e *data mining* na engenharia têxtil tem vindo a ganhar destaque, sendo amplamente aplicada na automatização de processos, análise preditiva e controlo de qualidade. Têm sido utilizados algoritmos como *Support Vector Machine* (Montesinos López et al., 2022b; Valkenborg et al., 2023), *Random Forest* (Hu &

Szymczak, 2023; Salman et al., 2024) e Redes Neurais Artificiais (Abdolrasol et al., 2021; Montesinos López et al., 2022a) para prever propriedades mecânicas e defeitos em tecidos de malha (Ahmed & Junayed, 2025).

Um estudo de Jebbor et al. (2024) demonstra que a Inteligência Artificial pode melhorar os processos de tomada de decisão nas cadeias de abastecimento através de capacidades preditivas e proativas. Utilizando um caso da indústria têxtil, é mostrada a aplicação do *machine learning* para prever disrupções na cadeia de abastecimento. O processo envolve a engenharia de variáveis preditivas, definição de métricas de desempenho adequadas ao caso de estudo e a análise sistemática dos algoritmos mais eficazes. A inteligência artificial pode reforçar os processos inteligentes de tomada de decisão nas cadeias de abastecimento, ao introduzir capacidades proativas e preditivas.

Já um estudo de Ersoz (2017) analisou dados de uma empresa produtora de calças para aplicar técnicas de *data mining* com o objetivo de prever a produção por trabalhador. Verificou-se que longas horas de trabalho, quer em horário normal quer em horário extraordinário, reduzem a produtividade. Para ajudar as fábricas têxteis a manterem-se competitivas, foram comparados vários modelos de previsão, nomeadamente Redes Neurais Artificiais, Regressão Linear Múltipla e os algoritmos CART (Abedinia & Seydi, 2024) e CHAID (McCarty & Hastak, 2007; Milanović & Stamenković, 2016), sendo avaliada a importância relativa de cada variável no desempenho da produção.

Na indústria têxtil existe uma grande necessidade de um sistema escalável de previsão da largura do tecido para resolver problemas relacionados com tecidos para os fabricantes de tecidos de malha. Um artigo de investigação de Sekhar et al. (2019), propõe um modelo escalável de previsão prioritária aproximada para prever a largura do tecido utilizando vários parâmetros têxteis. A previsão da largura do tecido é importante para os fabricantes de vestuário e de tecidos de malha, pois permite ajustar vários parâmetros para alcançar condições ótimas de produção. Também é útil para os produtores de fio, que podem assim fornecer o fio mais adequado para a malha.

Um estudo realizado por Makalesi et al. (2024), analisou fatores de risco para o controlo de acidentes de trabalho através de métodos de *data mining*, com base em dados de acidentes ocorridos em locais de trabalho do setor têxtil entre 2019 e 2021. O objetivo foi avaliar o desempenho de modelos de aprendizagem automática que pudessem servir de base para sistemas de previsão de acidentes. Para isso, foi criado um conjunto de dados específico com informações sobre acidentes no setor têxtil.

Após filtragem, foram selecionados os métodos de *machine learning* mais apropriados ao tipo de dados e ao objetivo do estudo. O estudo conclui que modelos de previsão de acidentes, baseados em *data mining*, podem ser eficazes na identificação de riscos significativos na indústria têxtil. Esses modelos permitem prever os tipos de lesões com base em características dos trabalhadores e dos processos de trabalho, contribuindo para a criação de mapas de risco e para a atuação proativa dos profissionais de Segurança e Saúde no Trabalho.

- **Medicina, Ciência e Engenharia** - Investigadores nestas áreas estão rapidamente a acumular dados importante que poderão permitir novos avanços. Por exemplo, para melhorar a compreensão do sistema climático da Terra, a NASA implementou uma série de satélites orbitais que estão a gerar permanentemente observações globais da superfície terrestre, dos oceanos e da atmosfera. No entanto, devido ao tamanho e à natureza espaço-temporal dos dados, os métodos tradicionais muitas vezes não são adequados para analisar esses conjuntos de dados. As técnicas de *data mining* podem ajudar os cientistas a responder a questões como: "Qual é a relação entre a frequência e a intensidade de distúrbios dos ecossistemas, como secas e furacões, com o aquecimento global?"; "De que forma a precipitação e a temperatura da superfície terrestre são afetadas pela temperatura da superfície oceânica?"; e "Quão bem podemos prever o início e o fim da estação de crescimento de uma região?"

Noutro exemplo, investigadores no campo da biologia molecular esperam utilizar grandes volumes de dados sobre o genoma para permitir uma melhor compreensão da estrutura e a função dos genes. No passado, os métodos tradicionais utilizados na área da biologia molecular permitiam apenas analisar alguns genes em cada estudo. Os recentes avanços ao nível da tecnologia de ferramentas biotecnológicas permitem comparar o comportamento de milhares de genes em diferentes situações. Estas comparações podem ajudar a determinar a função de cada gene e, possivelmente, isolar os genes responsáveis por determinadas doenças. No entanto, a natureza ruidosa e de alta dimensionalidade dos dados requer novos métodos de análise. Além de analisar dados de expressão genética, o *data mining* pode ser usado para enfrentar outros desafios biológicos importantes, como a previsão da estrutura de proteínas, alinhamento de sequências múltiplas, modelagem de vias bioquímicas e filogenética (Tan et al., 2019).

O crescente fardo das doenças crónicas tornou a sua prevenção e gestão uma prioridade global. Atualmente, quase um terço dos adultos em todo o mundo vive com mais de uma doença crónica. A utilização extensiva de tecnologia da informação na área da saúde levou à recolha de vastas quantidades de dados, ricos em

informações valiosas que ainda estão por explorar. O *big data* tem um enorme potencial para orientar decisões sobre trajetórias de saúde dos pacientes e a gestão de doenças crônicas (Chen et al., 2023).

- **Marketing** - Os dados são um elemento fundamental nos processos de tomada de decisão, permitindo que os sistemas de apoio à decisão resumam a informação de forma gráfica ou textual através de inteligência artificial avançada. Estes sistemas recolhem informações essenciais, incluindo dados sobre ativos, sistemas legados e dados relacionais, e realizam análises comparativas para prever tendências futuras em contextos específicos. A maioria das organizações integrou sistemas de apoio à decisão nas suas operações diárias. Uma prática comum nas organizações é recolher regularmente dados de vendas, juntamente com informações de orçamento e previsões, para ajustar as estratégias de *marketing*. Da mesma forma, os sistemas de apoio à decisão recolhem, analisam e reorganizam dados para facilitar decisões estratégicas (Kumar, 2020).

A estratégia de segmentação de mercado oferece diversos benefícios. O mais evidente é que os responsáveis pela tomada de decisões podem utilizar um *marketing mix* específico para alcançar um mercado mais pequeno com maior precisão. Esta abordagem permite uma utilização mais eficaz e eficiente dos recursos. Além disso, a segmentação de mercado reforça as relações entre os clientes e a empresa. Os resultados da segmentação podem também ser utilizados para definir estratégias competitivas específicas, como a diferenciação e a liderança em custos baixos (Huang et al., 2007). As redes sociais e o comércio eletrónico geram grandes volumes de dados passíveis de serem utilizados em investigação, o que provoca uma mudança significativa relativa à pesquisa sobre *big data* (dados não numéricos, como mensagens de texto) nas redes sociais, utilizando processamento de linguagem natural (NLP), destacando-se o grande impacto da influência do boca a boca (WoM) *online* de utilizadores e influenciadores (Sheth, 2021). Estudos demonstraram que a segmentação personalizada de clientes não só é mais lucrativa para as empresas, mas também que o conhecimento sobre o comportamento do cliente é um fator decisivo para o sucesso e o fracasso de um modelo de negócio (Alves Gomes & Meisen, 2023). Uma análise preditiva, com base em algoritmos de aprendizagem automática, permite prever o valor futuro dos clientes e a probabilidade de os manter. Diversas ferramentas de *data mining* são utilizadas para diferentes aplicações, conforme ilustrado na figura seguinte (Deldadehasl et al., 2025).

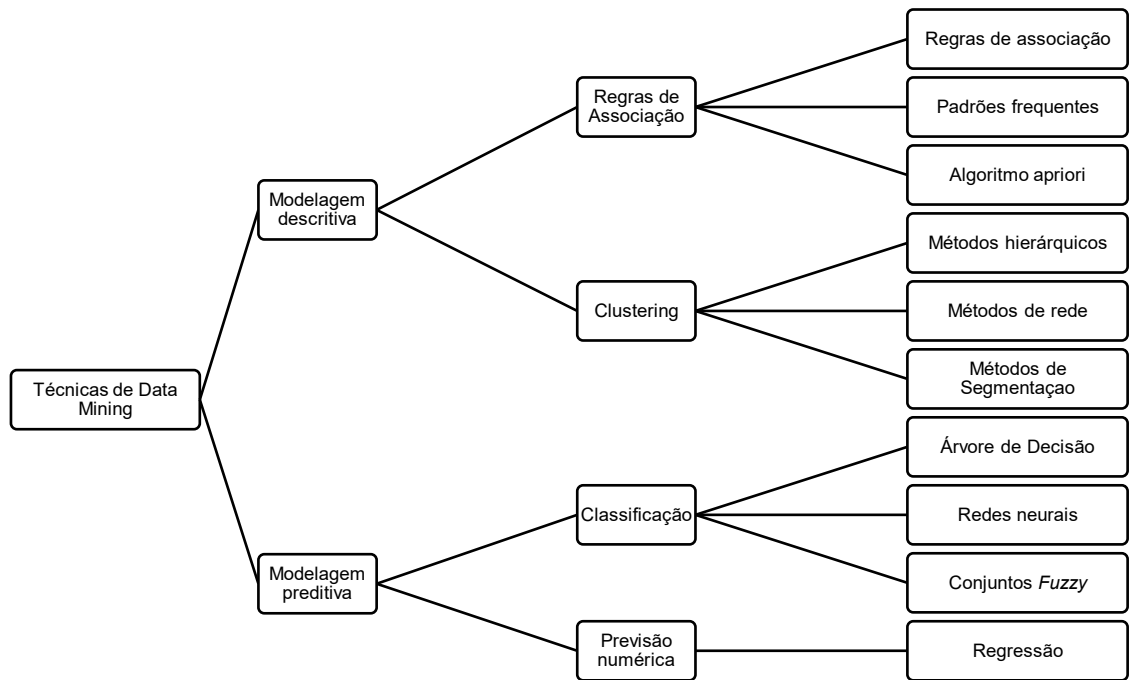


Figura 3 - Técnicas de Data Mining (Mahdiraji et al., 2019)

3. Métodos de *clustering* e de previsão

Este capítulo apresenta os métodos de *clustering* e de previsão considerados neste trabalho.

3.1. *Clustering*

O *machine learning* compreende diferentes temas, como análise regressiva, métodos de seleção de características e classificação. O tema da classificação envolve a atribuição de classes aos objetos de um conjunto de dados. Neste âmbito, existem três abordagens principais que podem ser consideradas para classificação: classificação supervisionada, semi-supervisionada e não supervisionada. No primeiro caso, as classes de alguns objetos são conhecidas antecipadamente, definindo o conjunto de treino, e um algoritmo é usado para obter os critérios de classificação. A classificação semi-supervisionada lida com o treino do algoritmo utilizando dados classificados e não classificados. Esta metodologia é comumente utilizada quando a classificação manual de um conjunto de dados se torna custosa. Por fim, a classificação não supervisionada, doravante referida como *clustering*, lida com a definição de classes a partir dos dados sem conhecimento da respetiva classificação (Rodriguez et al., 2019).

O *clustering* é uma ferramenta útil em ciência de dados, sendo um método que permite encontrar estruturas de *clusters* num conjunto de dados, caracterizado pela maior semelhança dentro do mesmo *cluster* e a maior dissimilaridade entre diferentes *clusters* (Sinaga & Yang, 2020). Os algoritmos de *clustering* são aplicados a uma variedade de tarefas, sem exclusão de campos específicos, desde que exista uma grande quantidade de dados disponíveis (Wegmann et al., 2021). A análise de *clustering* de dados, especialmente no campo do *data mining*, tem desempenhado um papel vital na organização e classificação apropriada dos dados ao longo do tempo. Simplificando, o *clustering* de dados ou análise de *clustering* é um processo que visa agrupar dados em classes diferentes, os *clusters*, de modo que objetos com atributos ou características semelhantes sejam agrupados na mesma classe ou *cluster* (Ezugwu et al., 2021), o que pode ser visualizado na seguinte Figura 4.

De acordo com Wegmann, Zipperling, et al. (2021), uma análise de *clustering* típica consiste em quatro etapas centrais com um caminho de *feedback*, conforme representado na Figura 5. A partir de um conjunto de dados, o primeiro passo é a seleção ou extração de características. Na seleção de características, são selecionadas características distintas de vários candidatos. Na extração de características, por outro lado, algumas transformações das características originais são usadas para criar características novas e significativas. Ambas são muito importantes para a eficácia das aplicações de *clustering* e uma seleção/extração de características adequadas pode reduzir significativamente a carga de trabalho e simplificar o

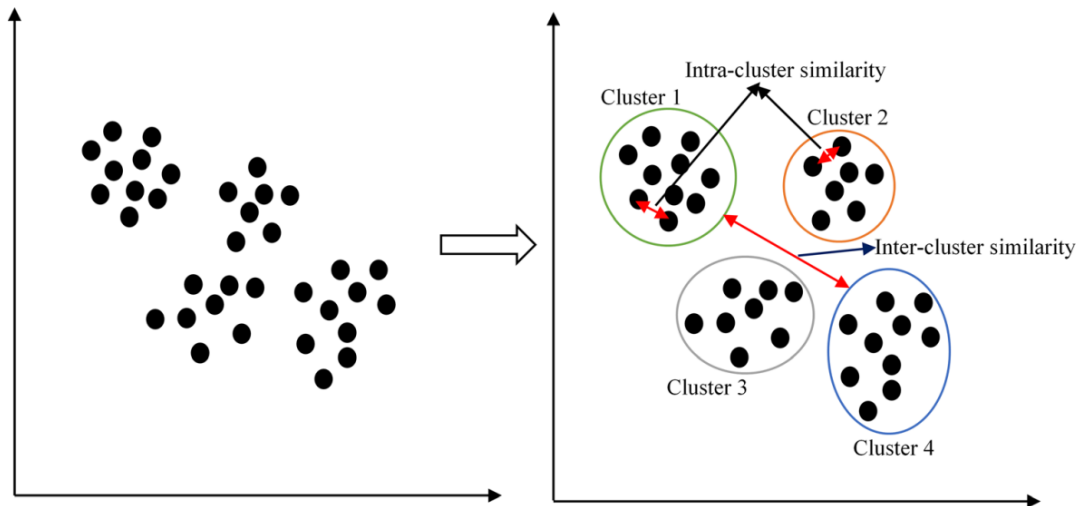


Figura 4 - Exemplo de clustering, representando similaridades intra e inter clusters(Ezugwu et al., 2021)

processo de *design* subsequente. Em geral, as características ideais devem ser imunes ao ruído, bem como fáceis de extrair e interpretar. O segundo passo é o desenvolvimento ou a seleção de um algoritmo de *clustering* eficaz. A seleção de um algoritmo de *clustering* afeta diretamente os resultados do *clustering*. Como o número de soluções possíveis disponíveis na literatura é muito vasto, torna-se muito importante estudar cuidadosamente as características do problema subjacente para selecionar ou projetar uma abordagem apropriada. Subsequentemente, a aplicação do algoritmo escolhido precisa ser validada. Dado um determinado conjunto de dados, um algoritmo de *clustering* encontrará uma divisão. A chave é dividir os dados de uma forma que contribua e permita a solução do problema inicial. Portanto, são necessários critérios de avaliação eficazes para os resultados do *clustering*. Num último passo, os resultados do *clustering* devem ser interpretados. O objetivo final desta fase é extrair informações significativas dos dados originais que resolvam o problema inicial de *clustering* de dados.

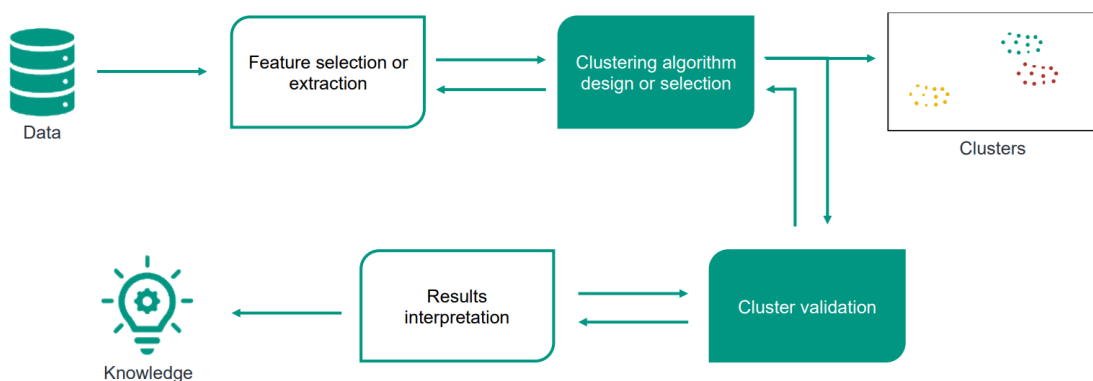


Figura 5 - Procedimentos de análise de clusters (Wegmann et al., 2021)

3.1.1. Algoritmos de *Clustering*

O objetivo do *clustering* de dados, também conhecido como análise de *clusters*, é descobrir agrupamentos naturais de um conjunto de padrões, pontos ou objetos. Uma definição operacional de *clustering* pode ser enunciada da seguinte forma: Dada uma representação de n objetos, encontrar K grupos com base numa medida de similaridade, de modo que as similaridades entre objetos no mesmo grupo sejam altas, enquanto as similaridades entre objetos em grupos diferentes sejam baixas (Jain, 2010).

De acordo com Ezugwu et al. (2021) os métodos de *clustering* podem ser amplamente classificados em duas categorias principais: *clustering* hierárquico ou partitivo. Os algoritmos de *clustering* hierárquico são procedimentos iterativos de *clustering*, que geram saídas semelhantes a uma árvore hierárquica ou dendrograma, mostrando uma sequência de *clustering* com cada um dos *clusters*. Em contraste, os algoritmos de *clustering* partitivo decompõe os conjuntos de dados em um conjunto de *clusters* disjuntos com base em critérios de otimização específicos. No método de *clustering* partitivo, encontramos os métodos de *clustering* tradicionais, como k-médias, *fuzzy c-means*. Os algoritmos de *clustering* k-médias e suas variantes, como k-medóides, dominam principalmente devido à sua simplicidade de implementação e sucesso empírico. A taxonomia dos algoritmos de *clustering* é apresentada na seguinte figura.

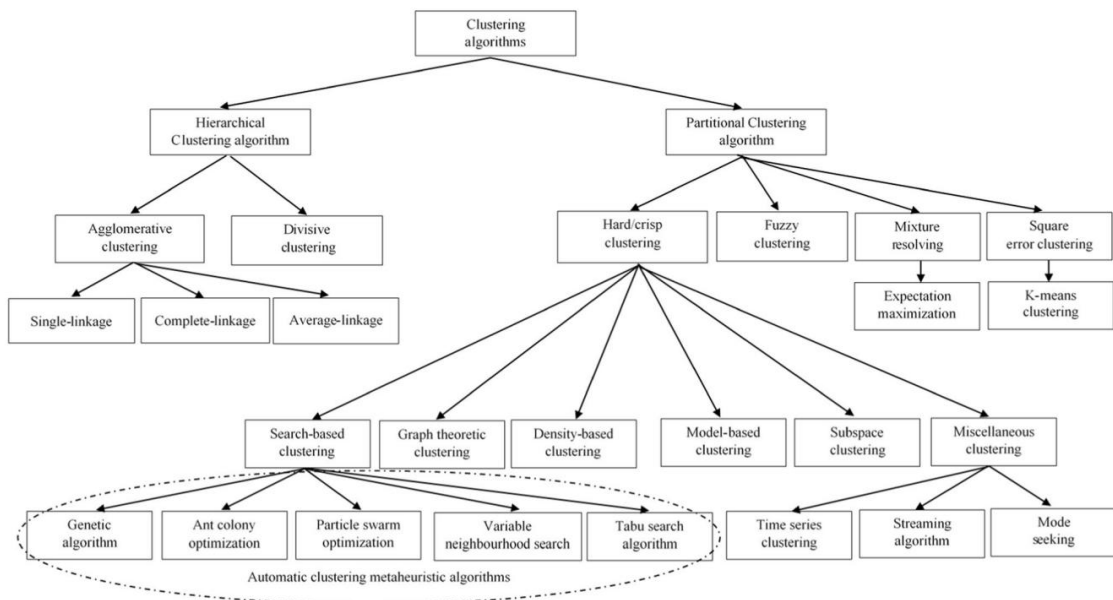


Figura 6 - Taxonomia dos algoritmos de *clustering* (Ezugwu et al., 2021)

A ideia básica destes algoritmos de *clustering* é considerar o centro dos pontos de dados como o centro do *cluster* correspondente. Os algoritmos k-médias e k-medóides atualizam o centro do *cluster*, por meio de computações iterativas, continuando o processo

até que algum critério de convergência seja atendido. O *K-medóides* é uma variante do K-médias projetada para lidar com dados discretos ou categóricos (Xu & Tian, 2015).

O k-médias define centróides, geralmente as médias de grupos de pontos, e é tipicamente aplicado a objetos em um espaço contínuo de n dimensões. O *k-medóides* define *medóides*, os pontos mais representativos de grupos de pontos, e pode ser aplicado a uma ampla gama de dados, contínuos ou não. Enquanto um centróide quase nunca corresponde a um ponto de dados real, um *medóide*, por definição, deve ser um ponto de dados real (Tan et al., 2019). Seguidamente é dada uma descrição mais pormenorizada de cada um dos algoritmos.

3.1.1.1. Algoritmo k-médias

O algoritmo k-médias é utilizado para agrupar dados em um número predefinido de *clusters*, representados por k centróides iniciais. Primeiro, cada ponto de dados é atribuído ao centróide mais próximo, formando *clusters* iniciais. Em seguida, os centróides são atualizados recalculando suas posições com base nos pontos atribuídos a cada *cluster*. Este processo de atribuição e atualização é repetido iterativamente até que não ocorram mudanças nos pontos atribuídos aos *clusters* ou, equivalentemente, até que os centróides se estabilizem. O objetivo é encontrar uma configuração de centróides que minimize a variância dentro dos *clusters*, resultando em uma partição dos dados onde pontos semelhantes estão agrupados juntos (Tan et al., 2019). Este algoritmo visa minimizar a soma do quadrado das distâncias e tem a seguinte expressão matemática:

$$SSE = \sum_{i=1}^K \sum_{X \in C_i} dist(c_i, X)^2$$

A Figura 7 seguinte ilustra a aplicação do algoritmo. Primeiro, são escolhidos k centróides iniciais, onde k é um parâmetro especificado pelo utilizador, ou seja, o número de *clusters* desejado. Na figura considera-se $k=3$. Cada ponto é então atribuído ao centróide mais próximo, e cada coleção de pontos atribuída a um centróide é um *cluster* (Figura 7 (a)). O centróide de cada *cluster* é então atualizado com base no cálculo da média dos pontos atribuídos ao *cluster* (Figura 7 (b)). Os passos de atribuição e atualização são repetidos (Figura 7 (b) a (d)) até que nenhum ponto mude de *cluster* ou, de forma equivalente, até que os centróides permaneçam os mesmos (Figura 7 (d)) (Tan et al., 2019).

A definição prévia do número de *clusters* é uma limitação do algoritmo k-médias. Isto deve-se ao facto de a classificação final poder depender fortemente da escolha do número de *clusters*. Além disso, o k-médias não é particularmente recomendado em casos onde os *clusters* não apresentam distribuição convexa ou têm tamanhos muito diferentes. Além do

mais, o algoritmo k-médias é sensível à seleção inicial dos centróides iniciais. Dadas estas limitações, muitas modificações deste algoritmo foram propostas, como o *k-medóides* e o k-médias++. No entanto, este algoritmo, além de ter um baixo custo computacional, pode fornecer bons resultados em muitas situações práticas, como na detecção de anomalias e segmentação de dados (Rodriguez et al., 2019).

É de notar que para atribuir um ponto ao centróide mais próximo, é necessária uma medida de proximidade que quantifique a noção de "mais próximo" para os dados específicos. A distância Euclidiana é usada para pontos no espaço Euclidiano, enquanto a similaridade cosseno é mais adequada para documentos. Outras medidas de proximidade, como a distância de Manhattan e a medida de Jaccard, também podem ser apropriadas dependendo do tipo de dados (Tan et al., 2019).

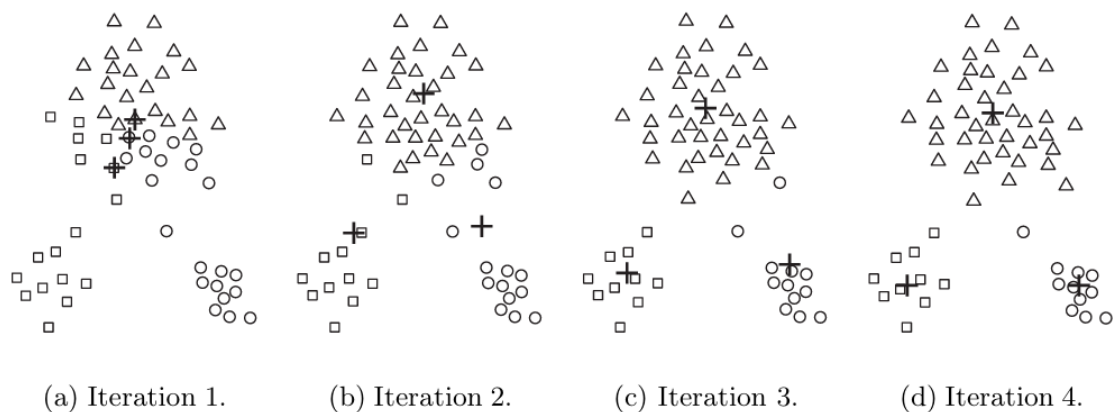


Figura 7 - Utilização do algoritmo k-médias para formação de três clusters cujos pontos são identificados por triângulos, círculos e quadrados (Tan et al., 2019)

3.1.1.2. Algoritmo k-medóides

O algoritmo k-médias tem como principal desvantagem a sua sensibilidade a *valores atípicos*, já que um valor muito alto ou muito baixo pode distorcer a distribuição dos dados. Em vez de utilizar a média dos valores dos objetos de um *cluster* como ponto de referência, pode-se utilizar um *medóide*, que é o objeto (pertencente ao *cluster*) mais centralmente localizado dentro do *cluster*. Desta forma, o método de partição pode ser realizado com base no princípio de minimizar a soma das dissimilaridades entre cada objeto e o seu ponto de referência. Este conceito é a base do método *k-medóides* (Kamal Kaur et al., 2014).

O algoritmo *K-medóides* é um método de formação de *clusters* usando os medóides, objetos representativos localizados no centro dos *clusters*. No algoritmo *K-medóides*, o processo de formação de k *clusters* é o seguinte. Inicialmente, selecionam-se aleatoriamente k-medóides e atribuem-se os restantes objetos aos *clusters* com o *medóide* mais próximo. Em

seguida, para cada *cluster*, o objeto com a menor média de distância para todos os objetos desse *cluster* é definido como o novo centro *medóide*. Os objetos restantes são então atribuídos aos *clusters* com o *medóide* mais próximo. Este processo é repetido até que os novos *medóides* e os *medóides* existentes sejam idênticos, ou seja, até que os *medóides* deixem de mudar. Este método representa uma melhoria em relação ao algoritmo k-médias, que é muito afetado por valores atípicos; de facto é menos sensível a valores atípicos, porque não utiliza a média dos objetos como objeto central (Lee et al., 2020)(Arora et al., 2016). A seguinte tabela apresenta um resumo das vantagens e desvantagens dos dois métodos:

	K-médias	K-medóides
Vantagens	• Bem conhecido e amplamente utilizado • Simples e fácil de usar	• Bem conhecido e amplamente utilizado • Menos sensível a valores atípicos • O tempo de execução por agrupamento é menor em comparação com o K-médias
Desvantagens	• O tempo de execução por agrupamento é maior em comparação com o K-medóides • É sensível a valores atípicos • O número de clusters é desconhecido	• O número de clusters é desconhecido • A complexidade é maior em comparação com o K-médias

Tabela 1 - Vantagens e desvantagens dos métodos de clustering (Herman et al., 2022)

Nesta dissertação, os objetos considerados são séries temporais e para medir a distância (dissemelhança) entre séries temporais é frequente utilizar as medidas de distância Euclidiana e *Dynamic Time Warping* (DTW) nos métodos de *clustering* K-médias e K-medóides (Li et al., 2022).

Os algoritmos baseados na distância Euclidiana comparam duas séries temporais ponto a ponto nos mesmos instantes de tempo, exigindo que ambas tenham o mesmo comprimento. Se as séries temporais a comparar tiverem comprimentos diferentes, é necessário ajustá-los, o que resulta inevitavelmente na perda de informação. Em contraste, os algoritmos baseados em DTW alinham as séries temporais de modo a minimizar a distância entre elas, permitindo a comparação de dados em momentos diferentes no tempo. Assim, ao utilizar DTW para comparar séries temporais de comprimentos eventualmente diferentes, é possível fazê-lo sem perda de dados, sendo bastante robusto face a distorções e deformações (Lee et al., 2020).

A figura seguinte ilustra a diferença entre utilizar a distância Euclidiana e a DTW.

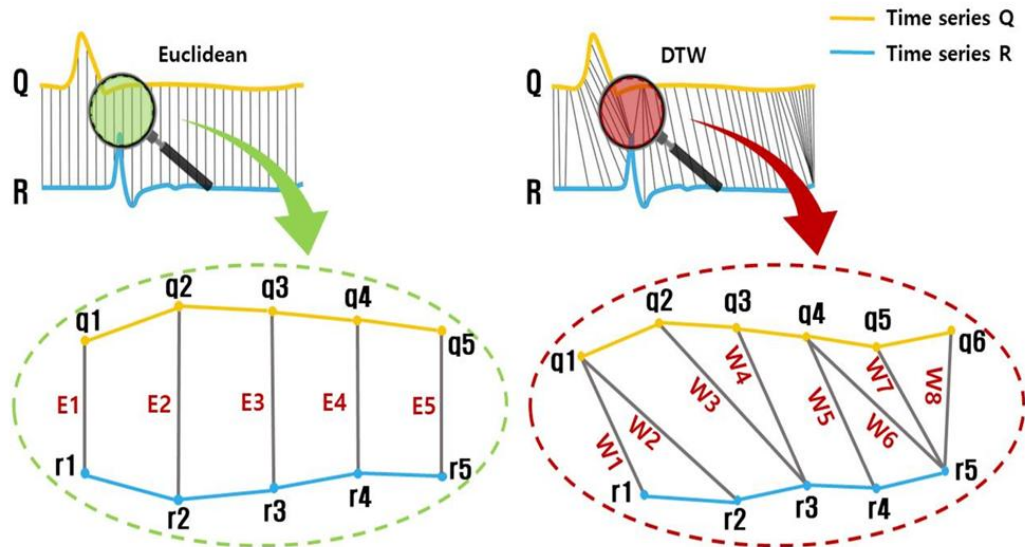


Figura 8 - Regras de alinhamento para o algoritmo Euclidiano e Dynamic Time Warping (Lee et al., 2020).

3.1.1.3. Dynamic Time Warping

O *Dynamic Time Warping* (DTW) é uma técnica amplamente conhecida para encontrar o alinhamento ideal entre duas sequências dependentes do tempo, sob certas restrições. Essencialmente, as sequências são ajustadas de forma não linear para corresponderem uma à outra. Inicialmente, o DTW foi utilizado para comparar diferentes padrões de voz no reconhecimento automático de fala. Em áreas como *data mining* e recuperação de informação, o DTW tem sido aplicado com êxito para lidar automaticamente com deformações temporais e variações de velocidade em dados dependentes do tempo (Muller, 2007). O alinhamento conseguido por aplicação desta técnica pode ser utilizado para encontrar regiões correspondentes ou para determinar a semelhança entre as duas séries temporais (Salvador & Chan, 2007).

Na figura seguinte cada linha vertical ou oblíqua conecta um ponto de uma série temporal ao ponto correspondente e similar na outra série temporal. As séries temporais apresentam valores semelhantes, mas foram separadas na figura para que as linhas verticais ou oblíquas entre elas possam ser visualizadas mais facilmente. Se as duas séries temporais fossem idênticas, todas as linhas seriam verticais, pois não seria necessária nenhuma distorção para "alinhar" as duas séries temporais (Salvador & Chan, 2007).

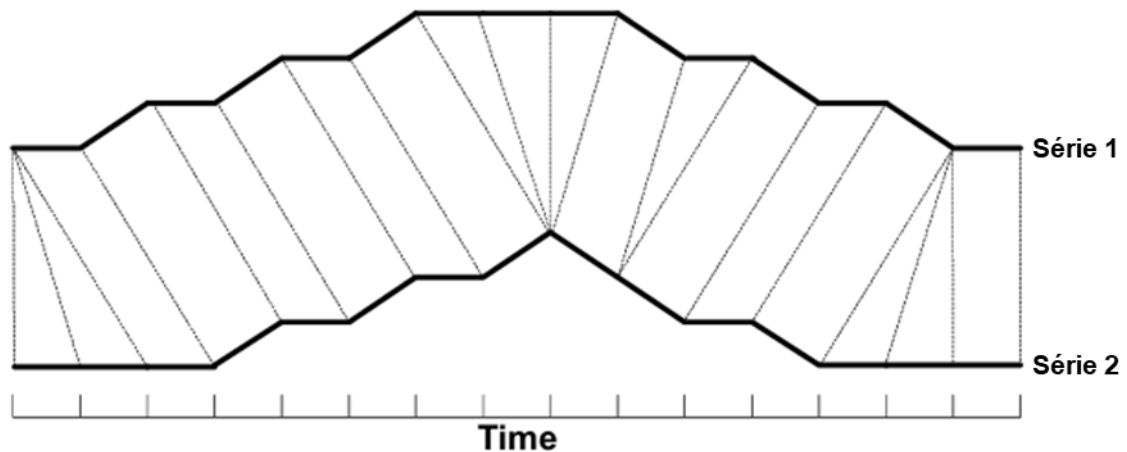


Figura 9 - Distorção (*warping*) entre duas séries temporais (Salvador & Chan, 2007)

A distância do percurso de distorção é a soma das distâncias entre os pontos correspondentes. Assim, duas séries temporais que são idênticas, exceto por um alongamento localizado do eixo temporal, terão distâncias de percurso de distorção de zero. O algoritmo DTW encontra o alinhamento ótimo entre duas séries temporais, mas possui uma complexidade que poderá limitar a sua utilidade a séries temporais relativamente pequenas (Salvador & Chan, 2007). O DTW pode medir a similaridade entre séries refletindo o atraso temporal causado pela distância entre os pontos de medição e, ao ser utilizado em algoritmos de *clustering* para agrupar séries temporais, poderá levar a melhores resultados (Lee et al., 2020). É de notar que as séries consideradas poderão ser transformadas antes de serem comparadas. Uma transformação comum é a estandardização, descrita na secção seguinte.

3.1.1.4. Estandarização

A estandardização dos valores de uma variável transforma esses valores: subtrai-lhes a média e divide a diferença pelo desvio-padrão. Assim, surge uma nova variável, cujos valores são o resultado da transformação mencionada. A média dos novos valores é 0 e o desvio padrão é 1.

Se diferentes variáveis forem consideradas juntas, como sucede no caso de séries temporais, onde a primeira variável corresponde ao que é registado no primeiro instante, a segunda no segundo instante, etc., a estandardização poderá favorecer os resultados, por exemplo, da aplicação de algoritmos de *clustering* ou agrupamento, por evitar que uma ou mais variáveis dominem outras pela simples razão de terem valores de magnitudes elevadas (Tan et al., 2019).

3.1.2. Coeficiente silhueta

Após um algoritmo de *clustering* processar um conjunto de dados, é crucial avaliar quão bem a partição obtida se ajusta a esses dados. Não existe um algoritmo de *clustering* ideal, pois diferentes algoritmos ou configurações dos mesmos produzem partições distintas, e nenhum ou nenhuma levam sempre aos melhores resultados, independentemente do problema considerado. Portanto, é necessário determinar várias partições e escolher a que melhor se adapta aos dados. Além disso, muitos algoritmos de *clustering* não determinam o número de *clusters* naturais e precisam dessa informação inicialmente (parâmetro *k*). Como essa informação raramente é conhecida, o algoritmo deve ser executado várias vezes com diferentes valores de *k*, e a melhor partição é então selecionada. O processo de estimar o quão bem uma partição se ajusta à estrutura subjacente dos dados é conhecida como validação de *clusters* (Arbelaitz et al., 2013).

A avaliação da qualidade dos resultados obtidos após aplicação de um algoritmo de *clustering* pode ser realizada através de vários métodos e dois dos mais populares são o método do cotovelo (*elbow method*) e o coeficiente silhueta. A eficácia do método do cotovelo depende das características do conjunto de dados. Quando o padrão do conjunto de dados é favorável, o método do cotovelo tende a funcionar bem. Em contrapartida, o coeficiente silhueta não depende dessas características, pois é um método baseado em distâncias: calcula-se a partir das distâncias médias entre os elementos dentro do mesmo *cluster* e o *cluster* mais próximo (Saputra et al., 2020).

O coeficiente silhueta é um método usado para avaliar a qualidade de agrupamentos na análise de *clusters*. A silhueta de um objeto é um valor entre -1 e 1 e mede o quão semelhante esse objeto é ao seu próprio *cluster* (similaridade) em comparação com outros *clusters* (dissimilaridade). O coeficiente silhueta é a média das silhuetas de todos os objetos e varia de -1 a 1, onde valores mais altos indicam uma melhor definição dos *clusters*.

Para obtenção do coeficiente silhueta, é necessário: a partição obtida (através da aplicação de alguma técnica de *clustering*) e a coleção de todas as proximidades (distâncias) entre os objetos (Rousseeuw, 1987).

O valor da silhueta $s(x_i)$ do objeto x_i , é definido da seguinte forma:

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max\{b(x_i), a(x_i)\}}$$

em que x_i é um objeto do *cluster* π_k , $a(x_i)$ é a distância média do objeto x_i a todos os outros objetos do *cluster* π_k e $b(x_i)$ é a menor distância média do objeto x_i aos outros objetos nos restantes *clusters* $\pi_l \neq \pi_k$, minimizada em relação a todos os *clusters* (ver Figura

10). Tem-se que $s(x_i)$ está entre -1 e 1. Uma silhueta próxima de 1 significa que o objeto está longe dos *clusters* vizinhos. Um valor próximo de 0 significa que o objeto está próximo da separação entre *clusters* vizinhos. Finalmente, um valor negativo significa que o objeto pode estar atribuído ao *cluster* errado. Assim, se a generalidade dos objetos tiver um valor da silhueta positivo e alto, então tal significa que, em geral, os objetos estão bem indicados aos *clusters* e que foram associados. Nesta situação, o valor do coeficiente silhueta s (média das silhuetas de todos os objetos) é positivo e alto e o *clustering* é apropriado, sendo o número de *clusters* adequado. Por outro lado, se muitos casos tiverem um valor da silhueta próximo de 0 ou negativo, então o valor do coeficiente silhueta s é próximo de zero ou negativo e o *clustering* não apropriado, por existirem ou muitos ou poucos *clusters*. Deste modo, o coeficiente de silhueta s pode ser usado para determinar o número ótimo de *clusters* num problema de *clustering*.

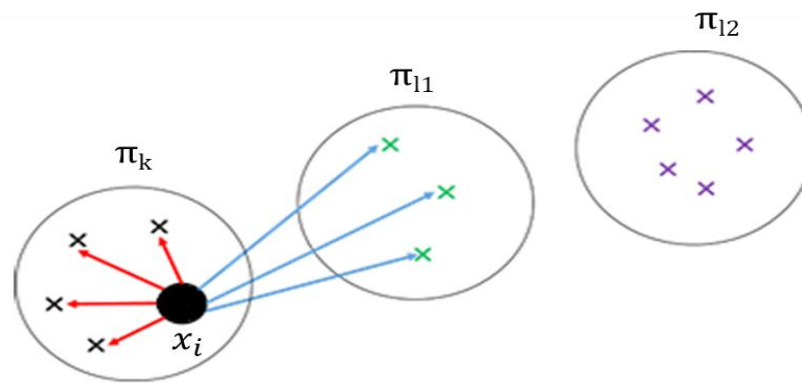


Figura 10 - Representação do processo de cálculo do coeficiente silhueta $s(x_i)$ para um determinado objeto x_i de um *cluster*. A vermelho estão representadas as distâncias para calcular $a(x_i)$ e a azul $b(x_i)$ (Shutaywi & Kachouie, 2021)

O cálculo do valor de $a(x_i)$ e $b(x_i)$, depende da fórmula utilizada no cálculo da distância. Assim, diferentes medidas de distância podem ser utilizadas, tais como a distância Manhattan, Euclidiana e a distância calculada através do DTW. No caso desta dissertação, os objetos são séries temporais. Se duas séries temporais forem idênticas, mas uma estiver ligeiramente deslocada ao longo do eixo do tempo, a distância Euclidiana considerará que elas são muito diferentes (Salvador & Chan, 2007). A vantagem do algoritmo DTW no cálculo da distância entre duas séries temporais reside na capacidade de comparar séries temporais de comprimentos desiguais e de incorporar funções de distância locais complexas nos passos intermediários (Zhang et al., 2021).

3.2. Previsão de séries temporais

A previsão é uma tarefa comum em ciência de dados que ajuda as organizações com o planeamento de capacidade, definição de metas e deteção de anomalias, entre outras

aplicações. Por exemplo, organizações de todos os setores da indústria precisam de fazer planeamento de capacidade para alocar eficientemente recursos escassos e de definir metas para medir o desempenho em relação a uma base de referência (Taylor & Letham, 2018). Uma vantagem competitiva sustentada é o objetivo final de todas as empresas, embora as opiniões sobre como alcançá-la tenham mudado ao longo do tempo. A necessidade de uma avaliação quantitativa das variáveis empresariais, e em particular das previsões futuras, mantém-se estreitamente ligada ao sucesso empresarial e à sua competitividade. Prever variáveis é um problema fundamental da investigação científica na economia empresarial (Kolkova, 2020). A previsão é necessária em muitas situações: decidir se deve construir outra central de produção de energia nos próximos cinco anos requer previsões da procura futura; agendar a equipa num *call center* na próxima semana requer previsões dos volumes de chamadas; manter um *stock* requer previsões das necessidades de inventário. As previsões podem ser necessárias com vários anos de antecedência (no caso de investimentos em capital) ou apenas alguns minutos antes (para o encaminhamento de telecomunicações). Quaisquer que sejam as circunstâncias ou horizontes temporais envolvidos, a previsão é uma ajuda importante para o planeamento eficaz e eficiente (Hyndman & Athanasopoulos, 2018). Consequentemente, previsões inadequadas podem resultar em um planeamento deficiente e em custos acrescidos para uma empresa. Como se deve proceder para fornecer as previsões trimestrais de vendas? Bom senso, intuição e uma consciência do estado da economia podem dar uma ideia geral ou "sensação" do que é provável que aconteça no futuro, mas converter essa sensação num número que possa ser usado como previsão de vendas para o próximo ano é difícil (D. Anderson et al., 2010). A previsão de procura é a estimativa das necessidades futuras de produtos ou serviços, baseada em dados presentes e passados, bem como em diferentes aspetos do mercado. Como todas as organizações enfrentam um futuro incerto e não se sabe qual será a procura de produtos no futuro, ao analisar os dados históricos e as informações atuais do mercado, pode-se prever a procura futura e fabricar os bens que terão maior procura num futuro próximo. Assim, é possível produzir antecipadamente os produtos necessários, de acordo com a procura do mercado. Desta forma, a previsão de procura é uma ferramenta essencial para um planeamento de gestão eficiente e eficaz. (Khan et al., 2020b).

Uma série temporal consiste numa sequência de observações de dados recolhidas em intervalos de tempo geralmente uniformes, como diariamente, mensalmente, anualmente, entre outros. A previsão de séries temporais é aplicada para estimar valores futuros a partir de dados históricos. Sendo importante, a previsão é também desafiante, em particular quando os dados correspondem a séries temporais não estacionárias, ou seja, que não mantêm as mesmas características ao longo do tempo. A finalidade desta previsão é detetar padrões,

tendências ou ciclos inerentes aos dados e utilizá-los como base para a extrapolação de valores futuros (Deb et al., 2017).

Os métodos de previsão dividem-se em qualitativos e quantitativos. Os qualitativos usam julgamento de especialistas e são adequados quando os dados históricos não estão disponíveis. Já os quantitativos são aplicáveis quando há dados passados quantificáveis e quando se pode assumir que padrões anteriores se manterão. Se os dados históricos se restringem aos valores passados da variável a ser prevista, o procedimento de previsão é denominado método de série temporal e os dados históricos são referidos como uma série temporal. O objetivo da análise de séries temporais é descobrir um padrão nos dados históricos ou na série temporal e, em seguida, extrapolar esse padrão para o futuro; a previsão baseia-se exclusivamente em valores passados da variável e/ou em erros de previsões passadas (D. Anderson et al., 2010).

Conforme já foi referido, uma série temporal é uma sequência ordenada de valores registados a intervalos iguais de tempo. A análise de séries temporais pode ser dividida em duas partes. A primeira parte consiste em obter a estrutura e o padrão subjacente dos dados observados. A segunda parte envolve ajustar um modelo para fazer previsões futuras (Deb et al., 2017). A análise de séries temporais divide-se em duas categorias: univariada, que lida com uma única variável ao longo do tempo, e multivariada, que considera múltiplas variáveis e suas interações. Uma abordagem comum na análise de séries temporais é decompor a série nas seguintes três componentes:

Tendência – O movimento geral que a variável exibe durante o período de observação, sem considerar a sazonalidade e irregularidades.

Sazonalidade – As flutuações periódicas da variável em análise, consistindo em efeitos que são estáveis em termos de tempo, magnitude e direção.

Resíduo – A parte restante, geralmente inexplorável, da série temporal. Estes resíduos podem, às vezes, ser suficientemente altos para mascarar a tendência e a sazonalidade (Deb et al., 2017).

Os métodos de previsão causal baseiam-se na suposição de que a variável que se está a prever tem uma relação de causa e efeito com uma ou mais variáveis. Ao considerar a análise de regressão como uma ferramenta de previsão, pode-se ver o valor da série temporal que se quer prever como a variável dependente. Assim, se se conseguir identificar um bom conjunto de variáveis independentes ou explicativas relacionadas, pode-se desenvolver uma equação de regressão estimada para prever ou estimar a série temporal. Ao tratar o tempo como a variável independente e a série temporal como a variável dependente, a análise de regressão também pode ser usada como um método de série temporal (D. Anderson et al., 2010).

3.2.1. Padrões de séries temporais

Conforme já foi referido, uma série temporal é uma sequência de observações de uma variável medida em pontos sucessivos no tempo ou ao longo de períodos sucessivos. O padrão dos dados é crucial para entender o comportamento passado da série temporal e, se houver indícios desse comportamento continuar no futuro, pode-se usar esse padrão para escolher um método de previsão adequado. Para identificar o padrão subjacente nos dados, um primeiro passo útil é construir um gráfico da série temporal, conhecido por cronograma. Um cronograma apresenta graficamente a relação entre o tempo (eixo horizontal) e a variável da série temporal (eixo vertical). Muitas vezes, é possível identificar padrões comuns nos dados ao examinar esse gráfico, como um padrão horizontal, com tendência, sazonal, com tendência e sazonalidade ou cíclico. (D. Anderson et al., 2010). Estes padrões são descritos de seguida.

3.2.1.1. Padrão horizontal

Uma série com padrão horizontal verifica-se quando os valores registados flutuam em redor de uma média constante (ver Figura 11). Uma série temporal estacionária é uma série cujas propriedades estatísticas são independentes do tempo, significando que o processo de geração de dados tem uma média e uma variabilidade constantes ao longo do tempo. Assim o cronograma de uma série temporal estacionária exibirá um padrão horizontal. No entanto, um padrão horizontal poderá não ser evidência suficiente de estarmos na presença de uma série estacionária.

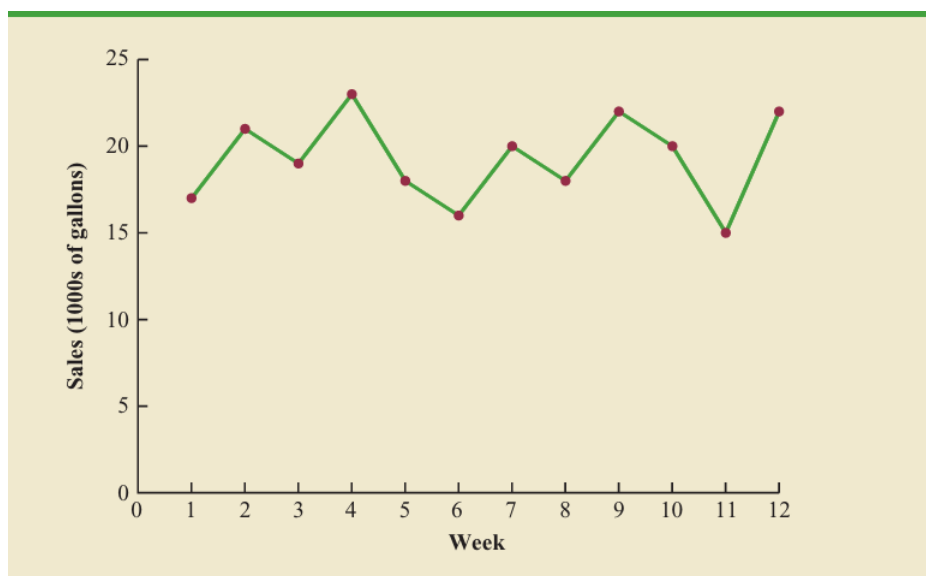


Figura 11 - Exemplo de uma série temporal de padrão horizontal (D. Anderson et al., 2010)

3.2.1.2. Padrão com tendência

Os dados de séries temporais geralmente apresentam flutuações aleatórias. No entanto, também podem mostrar mudanças graduais para valores mais altos ou mais baixos ao longo do tempo, o que caracteriza um padrão de tendência (ver Figura 12). Algumas tendências podem seguir padrões não lineares, como o crescimento exponencial, em que a taxa de mudança não é constante, mas aumenta de forma percentual.

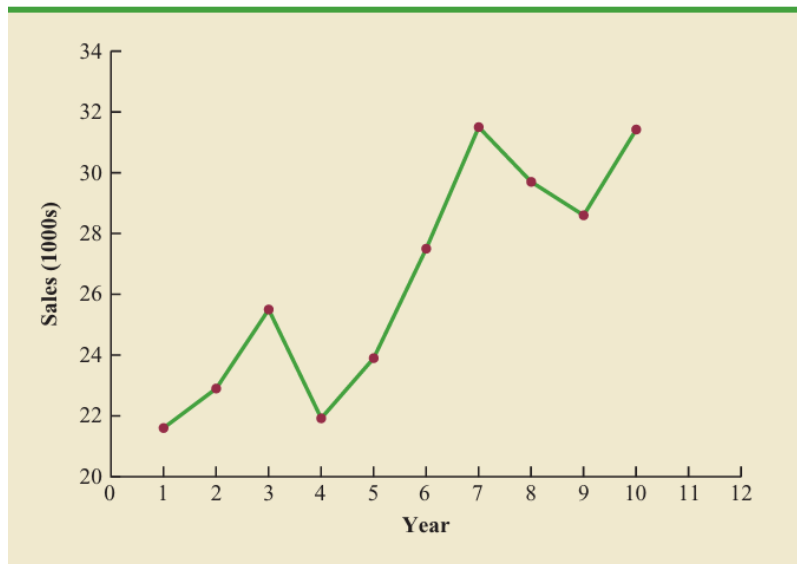


Figura 12 - Exemplo de uma série temporal com padrão de tendência (crescente) (D. Anderson et al., 2010)

3.2.1.3. Padrão sazonal

A tendência de uma série temporal pode ser identificada pela análise dos movimentos de vários anos nos dados históricos. Padrões sazonais são reconhecidos através da repetição de padrões em períodos sucessivos, como anos (ver Figura 13).

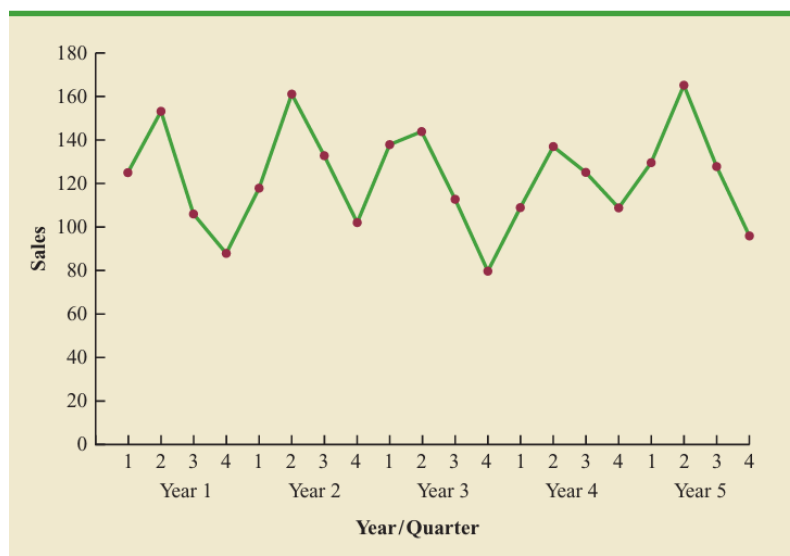


Figura 13 - Exemplo de uma série temporal com padrão sazonal (D. Anderson et al., 2010)

3.2.1.4. Padrão com tendência e sazonalidade

Algumas séries temporais combinam tendências e padrões sazonais (ver Figura 14). Para estas situações, é necessário utilizar um método de previsão que consiga lidar tanto com a tendência quanto com a sazonalidade.



Figura 14 - Exemplo de uma série temporal com padrão com tendência e sazonalidade (D. Anderson et al., 2010)

3.2.1.5. Padrão cíclico

Um padrão cíclico existe se o gráfico de uma série temporal mostrar uma sequência alternada de pontos abaixo e acima da linha de tendência, durando mais de um ano. Muitas séries temporais económicas exibem comportamento cíclico com períodos regulares de observações abaixo e acima da linha de tendência. Os efeitos cíclicos são frequentemente combinados com os efeitos de tendência de longo prazo e são chamados de efeitos de ciclo de tendência (ver Figura 15).

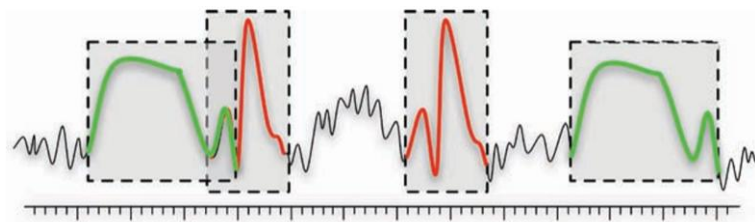


Figura 15 - Exemplo de uma série temporal com padrão cíclico (Esling & Agon, 2012)

3.2.2. Métodos de previsão

O padrão subjacente de uma série temporal é crucial na escolha do método de previsão. Um cronograma deve ser um dos primeiros passos para procurar determinar o método adequado. Por exemplo, se houver um padrão horizontal, deve-se selecionar um

método apropriado para esse tipo de padrão, sendo o método da média, descrito adiante, o mais simples.

Alguns métodos de previsão são, de facto, extremamente simples e surpreendentemente eficazes (Hyndman & Athanasopoulos, 2018). De seguida, são apresentados alguns destes métodos mais simples de previsão de séries temporais e, mais adiante, métodos mais elaborados, de regressão e autorregressão.

3.2.2.1. Método da média (Mean)

Neste método, as previsões de todos os valores futuros são iguais à média aritmética dos dados históricos. Se denotarmos os dados históricos por $y_1, \dots, y_T$ pode-se escrever a previsão para o tempo $T+h$ como:

$$\hat{y}_{T+h|T} = \bar{y} = \frac{(y_1 + \dots + y_T)}{T}$$

3.2.2.2. Método ingénuo (Naive)

Este método funciona destacadamente bem para variadas previsões económicas e financeiras. No método ingénuo, definem-se todas as previsões como o valor da última observação, ou seja,

$$\hat{y}_{T+h|T} = y_T$$

3.2.2.3. Método ingénuo sazonal (SNaive)

Este método é semelhante ao anterior e é útil para dados altamente sazonais. Neste caso, cada previsão é definida como o último valor observado da mesma época (por exemplo, o mesmo mês do ano anterior). Formalmente, a previsão para o tempo $T + h$ é:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$$

em que m é o período sazonal e k é a parte inteira de $\frac{(h-1)}{m}$. A título de exemplo, com dados mensais, a previsão dos valores dos meses de fevereiro é igual ao valor registado para o último mês de fevereiro observado.

3.2.2.4. Método de deriva (Drift)

Uma variação do método ingênuo é permitir que as previsões aumentem ou diminuam ao longo do tempo, onde a quantidade de mudança ao longo do tempo (chamada de "drift" ou "deriva") é definida como a mudança média observada nos dados históricos. Assim, a previsão para o tempo $T+h$ é dada por:

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1}) = y_T + h \left(\frac{y_T - y_1}{T-1} \right)$$

Este método equivale a desenhar uma linha entre a primeira e a última observação e a extrapolar as previsões para o futuro.

A figura seguinte ilustra a aplicação de parte dos métodos descritos,

Google stock (daily ending 6 Dec 2013)

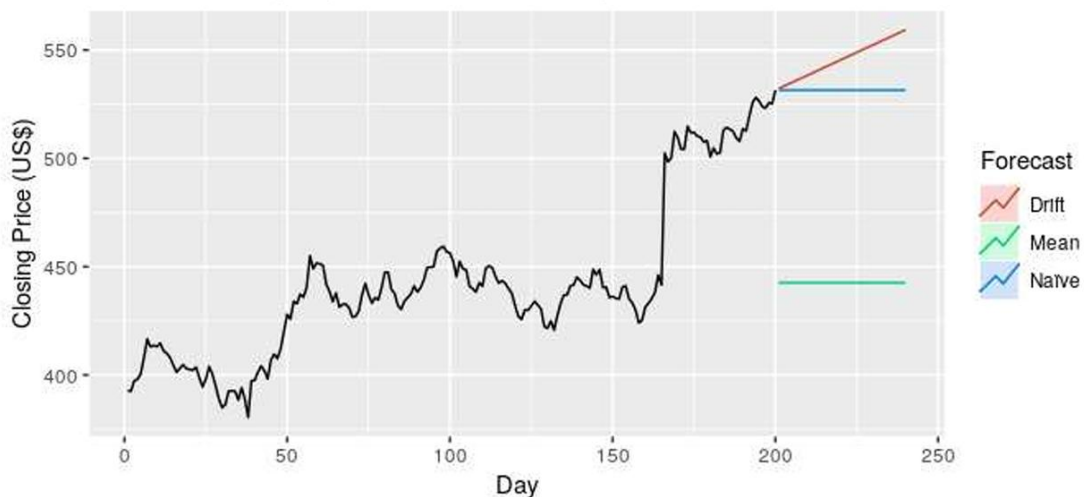


Figura 16 - Exemplo dos três métodos de previsão assinalados (Hyndman & Athanasopoulos, 2018)

3.2.2.5. Análise de regressão

A análise de regressão é uma técnica estatística usada para modelar e investigar as relações entre uma variável de resposta e uma ou mais variáveis preditivas. O objetivo final de um estudo de análise de regressão é gerar um modelo que possa ser utilizado para prever valores futuros da variável de resposta, dados os valores das variáveis preditivas (Montgomery et al., 2015).

O modelo de regressão linear simples envolve apenas uma variável preditiva e é expresso da seguinte forma:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

em que y é a variável de resposta, x é a variável preditiva, β_0 e β_1 são coeficientes e ε é um termo de erro.

Os parâmetros do modelo ou coeficientes de regressão β_0 e β_1 têm as seguintes interpretações: β_1 é a inclinação da reta e β_0 é a interceção da mesma com o eixo dos yy . A inclinação β_1 mede a variação na média da variável de resposta y para uma mudança unitária na variável preditiva x . Estes parâmetros são habitualmente desconhecidos e precisam ser estimados a partir de uma amostra de dados (x_i, y_i) para $i = 1, 2, \dots, n$. O termo de erro ε explica as diferenças entre os dados reais e a reta definida pela equação do modelo. Considera-se ε como um erro estatístico, geralmente assumido como uma variável aleatória normalmente distribuída com média zero e variância constante σ^2 .

Os modelos de regressão incluem muitas vezes mais do que uma variável preditiva. Neste caso, dizem-se modelos de regressão múltipla. Se existirem k variáveis preditivas $x_1, x_2, \dots, x_k$, o modelo de regressão linear múltipla é dado por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Os parâmetros $\beta_0, \beta_1, \dots, \beta_k$ neste modelo são chamados de coeficientes de regressão parcial, pois indicam o efeito de cada preditor na variável y , assumindo que todos os outros preditores no modelo permanecem constantes.

Se a variável de resposta y e as preditivas $x_1, x_2, \dots, x_k$ variarem com o tempo t , como é o caso nesta dissertação, em que os dados são séries temporais, o modelo de regressão escreve-se da seguinte forma:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, t = 1, 2, \dots, T, \dots$$

Os coeficientes $\beta_0, \beta_1, \dots, \beta_k$ numa regressão linear são normalmente estimados por aplicação do método dos mínimos quadrados a partir de uma amostra de dados $(x_{1i}, x_{2i}, \dots, x_{ki}, y_i)$ para $i = 1, 2, \dots, n$. O objetivo do método é encontrar as estimativas $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ que minimizam a soma dos quadrados dos resíduos, definidos como a diferença entre os valores observados y_i e os valores previstos $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_k x_{ki}$. Assim, o método dos mínimos quadrados resolve o seguinte problema:

$$\min \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Para tal pode-se recorrer ao *software* Microsoft Excel (Buzrul, 2024; Yan, 2022) ou a outros, como o R (Kamel & Abonazel, 2023).

3.2.2.6. ARIMA (Autoregressive Integrated Moving Average)

O modelo autorregressivo de média móvel (ARMA) é um método importante para o estudo e previsão de séries temporais. O modelo autorregressivo integrado de média móvel (ARIMA) é baseado no modelo ARMA. A diferença é que o modelo ARIMA converte dados não estacionários em dados estacionários antes de trabalhar com eles (Mondal et al., 2014). A transformação da série temporal em estacionária é conseguida por aplicação de um processo de diferenciação que transforma a série y numa série diferenciada y' . Recorde-se que uma série temporal pode ser considerada estacionária se suas propriedades estatísticas forem constantes ao longo do tempo. Assim, a equação ARIMA para uma série temporal é uma equação linear em que a entrada consiste no desfasamento da variável dependente y' , juntamente com o desfasamento do erro de previsão (Deb et al., 2017). Em palavras, o modelo genérico é o seguinte:

$$\text{Resultado } y' = \text{constante} + \text{soma ponderada de um ou mais valores passados de } y' \\ + \text{soma ponderada de um ou mais valores passados do erro } \varepsilon$$

Os termos desfasados da série temporal estacionária são chamados de "autorregressivos", enquanto os termos desfasados dos erros são chamados de "média móvel". Uma série temporal que precisa de ser diferenciada para se tornar estacionária é considerada uma versão "integrada" de uma série estacionária. Os modelos ARIMA são denotados como ARIMA (p,d,q), onde 'p' representa a ordem da parte autorregressiva, 'd' denota o grau de diferenciação necessária para tornar a série estacionária e 'q' representa a ordem da parte de média móvel (Deb et al., 2017).

Neste contexto, um modelo ARIMA pode ser formalmente escrito da seguinte forma (Hyndman & Athanasopoulos, 2018):

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Os modelos ARIMA também são capazes de modelar dados sazonais. Um modelo ARIMA sazonal é formado ao incluir os termos sazonais no modelo ARIMA visto antes e é denotado por:

$$\text{ARIMA } \underbrace{(p, d, q)}_{\text{Parte não sazonal do modelo}} \underbrace{(P, D, Q)}_{\text{Parte sazonal do modelo}} m$$

em que m é o número de observações por ano. Note-se que são usadas letras minúsculas para descrever a parte não sazonal e maiúsculas para descrever a parte sazonal. A parte sazonal do modelo consiste em termos que são similares às componentes não sazonais, mas envolvem retrocessos do período sazonal (Hyndman & Athanasopoulos, 2018).

O *software R* é uma ferramenta amplamente utilizada para a modelação de séries temporais, especialmente na aplicação de modelos ARIMA. Com a ajuda da biblioteca **forecast**, o R permite não apenas ajustar modelos ARIMA, mas também automatizar a escolha dos parâmetros p , P (autoregressivo), d , D (diferenciação) e q , Q (média móvel), por meio da função **auto.arima()**.

Além do ajuste automático, o R oferece ferramentas avançadas para a validação e diagnóstico do modelo que permitem verificar a qualidade do ajuste, analisando os resíduos do modelo para garantir que eles sejam aleatórios e não apresentem padrões, confirmando assim a adequação do modelo.

3.2.2.7. Avaliação dos resultados de previsão

A aplicação dos diferentes modelos de previsão resulta em diferentes previsões do valor da variável para cada um dos períodos da série temporal. Assim, para que se possa avaliar a qualidade das previsões obtidas por aplicação de cada um dos modelos, é necessário calcular os erros das mesmas. Para tal, recorre-se a métricas usuais, como o *Mean Absolute Error* (MAE) e o *Mean Absolute Percentage Error* (MAPE). Ambas são baseadas na noção de erro de previsão, que é a diferença $y_i - \hat{y}_i$ entre o valor observado da variável, y_i , e o valor previsto pelo modelo em consideração, $\hat{y}_i$ (Chicco et al., 2021). As duas métricas referidas anteriormente estão definidas da seguinte forma:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

em que n é o número de observações consideradas.

Assim, o MAE é a média dos desvios absolutos entre os valores observados da série e as correspondentes previsões fornecidas pelo modelo em avaliação. Note-se que o MAE é expresso nas mesmas unidades dos valores da série. Por seu turno, o MAPE é a média dos

desvios relativos entre os valores observados da série e as correspondentes previsões. O MAPE não tem unidades e é expresso em percentagem. Finalmente é de relevar que o MAE pondera todos os desvios $|y_i - \hat{y}_i|$ da mesma forma, mas o MAPE não. De facto, no primeiro caso a ponderação é sempre igual a 1 e no segundo caso é igual a $\frac{1}{|y_i|}$. Deste modo, o MAE dá a mesma importância a todos os desvios e o MAPE dá mais importância a desvios em que o valor observado da série tenha menor magnitude.

4. Lavandaria Industrial

Uma lavandaria industrial é uma empresa do sector têxtil que presta serviços de lavagem e acabamento têxtil, inserindo-se no processo produtivo industrial. Estes serviços caracterizam-se por processos de lavagem, secagem, tratamento, dobragem e transporte de produtos têxteis.

A lavandaria em análise, Precioso & Morais, Lda. é uma microempresa localizada em Braga e oferece uma variedade de tipos de lavagem e tratamentos, sendo os mesmos aplicados a diferentes tipos de peças.



Figura 17 - Imagens da lavandaria industrial Precioso & Morais

Esta empresa nascida em 1997, conta 15 trabalhadores e presta serviços para empresas do sector têxtil da região do Cávado, com alguns clientes inseridos noutras regiões industriais. O CAE da empresa é o 96010 - Lavagem e limpeza a seco de têxteis e peles, restringindo-se à lavagem de “limpos”, não disponibilizando o serviço de lavagem de “sujos”, nomeadamente roupa de hotelaria ou similares.

A lavandaria tem mantido nos últimos anos um plano de investimento na modernização dos equipamentos, com a aquisição de novas máquinas de lavar e secadores, bem como um avultado investimento na área da eficiência energética. Neste capítulo, destacam-se a aposta nas energias renováveis, com uma forte componente de energia solar, bem como, na produção de vapor através de fontes renováveis, nomeadamente biomassa, estando, neste momento, cada vez mais independente de combustíveis fósseis.

Em termos sociais, a empresa tem uma política de abertura, empregando diferentes nacionalidades de trabalhadores, dando oportunidades a trabalhadores estudantes de conciliarem os seus horários laborais e de estudo e ainda apoiando diversas causas sociais, destacando-se o apoio ao Banco Alimentar Contra a Fome, em que em todas as campanhas cede as carrinhas de transporte para apoio ao transporte de alimentos.

A lavandaria industrial, em termos processuais, emite para cada serviço prestado uma Guia de Transporte, processada por um *software* de faturação, onde estão registados vários dados relativos ao serviço prestado, nomeadamente:

- Data de Emissão;
- Tipo de Lavagem (existem 20 tipos);
- Tipo de Peça (existem 43 tipos);
- Quantidade;
- Preço Unitário;
- Número do cliente (existem 95 clientes);
- Total Líquido.

A lavandaria em estudo não é uma empresa orientada por dados. Apesar de realizar a recolha de dados, não é feito processamento, nem a análise, dos mesmos. Assim, os processos de decisão que ocorrem ao nível da gerência da empresa não são apoiados em dados.

Com este estudo, pretende-se dar um primeiro passo no sentido de tornar esta empresa numa organização orientada por dados. Para tal, pretende-se contribuir para duas das sete características definidas por (Assur & Rowshankish, 2022) e indicadas no capítulo anterior, nomeadamente para a **integração** e para o **processamento** de dados. De igual forma se pretende iniciar um processo de transformação com a **definição de uma estratégia** para a utilização futura de dados, de forma a integrar todo o sistema no processo de tomada de decisão sobre o modelo de negócio, seja ao nível de investimentos futuros em equipamentos, seja na segmentação dos clientes, seja numa estratégia de *marketing*.

A lavandaria sobre a qual este trabalho incide não faz qualquer análise dos dados armazenados no *software* mencionado anteriormente, sendo que os milhões de registos individuais são apenas utilizados para efeitos de faturação. Neste sentido, para permitir uma

melhor compreensão da atividade empresarial nos últimos anos, foram recolhidos os dados de faturação desta organização, respeitantes aos anos 2018 a 2023.

O tratamento e a análise destes dados, correspondentes a séries temporais, permitem a obtenção de informações importantes sobre a operação recente desta empresa, possibilitando que a informação recolhida permita uma extrapolação sobre o desenvolvimento do modelo de negócio. Isto reveste-se de particular importância, nomeadamente ao nível da previsão de evolução da receita, no sentido de preparar e precaver as necessidades futuras que garantam a sustentabilidade e o crescimento da empresa em análise.

Os dados recolhidos sobre a operação da lavandaria no período de seis anos, de 2018 a 2023, têm uma dimensão considerável, totalizando cerca de 5 milhões de registos individuais. Assim, para que seja possível uma análise adequada dos mesmos, torna-se conveniente o agrupamento dos dados em conjuntos de características similares. Para tal foi utilizada a ferramenta de *clustering*, descrita genericamente na secção 3.1.

A lavandaria industrial oferece uma variedade de tipos de lavagem e tratamentos, aplicados a diferentes tipos de peças. Com 20 tipos de lavagem e 43 tipos de peças, combinados com uma quantidade significativa de dados ao longo de seis anos, o *clustering* pode ajudar a identificar padrões comuns de serviços prestados e com isso determinar estratégias de *marketing* e otimização de processos.

Com 95 clientes diferentes, a ferramenta de *clustering* pode agrupar clientes com base nos seus comportamentos de compra, tipos de serviços utilizados e frequência de pedidos. Isso pode ajudar a identificar segmentos de mercado, como clientes que geram mais receita ou que utilizam serviços específicos. A empresa pode usar essas informações para personalizar ofertas, melhorar a fidelização de clientes e desenvolver estratégias de vendas direcionadas.

A utilização de dados históricos agrupados pela ferramenta de *clustering* pode ajudar na previsão de procura futura. Por exemplo, podem identificar-se padrões sazonais ou tendências de crescimento em certos tipos de serviços. Essas previsões são essenciais para a preparação de recursos e para garantir que a empresa possa atender à procura, sem comprometer a qualidade dos serviços.

A utilização da ferramenta de *clustering* para analisar os dados da lavandaria industrial é fundamental para transformar grandes volumes de dados brutos em informações úteis. Ao agrupar dados com características similares, a empresa pode compreender melhor os processos, os clientes e o desempenho. Isto facilita a tomada de decisões estratégicas, mas também apoia o desenvolvimento sustentável e o crescimento da empresa. A aplicação do *clustering* na análise dos dados da lavandaria industrial exemplifica como técnicas

avançadas de análise de dados e computação podem ser poderosas ferramentas para otimizar operações e impulsionar o sucesso empresarial.

Assim, sabendo que a lavandaria em questão ainda não é orientada por dados, uma vez que, apesar de recolher informações, não realiza o seu processamento ou análise, nem adota uma gestão baseada em conhecimento extraído dessas informações, este estudo propõe um primeiro passo para transformar a empresa, promovendo a integração e processamento de dados. A análise dos dados de faturação entre 2018 e 2023 visa fornecer informações importantes para futuramente poder vir a apoiar decisões estratégicas e operacionais, contribuindo para o desenvolvimento sustentável do modelo de negócio. A adoção de uma estratégia orientada por dados permitirá otimizar investimentos, segmentação de clientes e *marketing*, assegurando o crescimento futuro da empresa.

5. Metodologia

A metodologia desenvolvida nesta dissertação passou pela aplicação de ferramentas de análise de dados, a dados recolhidos da atividade de uma lavandaria industrial no período compreendido entre o ano 2018 e 2023. Numa fase inicial os dados foram recolhidos, seguidamente foram organizados, seguindo-se uma fase de análise e agrupamento dos mesmos e de aplicação de modelos de previsão. Este capítulo pormenoriza a metodologia seguida nas fases referidas.

5.1. Recolha e processamento de dados

Para elaboração deste estudo, foram recolhidos dados relativos à faturação da lavandaria industrial, descrita no capítulo anterior, relativos a um período de seis anos, nomeadamente entre 2018 e 2023. Os dados recolhidos através do *software* de faturação usado pela empresa permitem avaliar o desenvolvimento do negócio, nomeadamente a produção e a faturação.

A empresa presta um serviço e, como tal, no final da prestação do mesmo, emite uma guia, que acompanha a mercadoria no seu transporte até ao cliente. Cada uma destas guias de transporte tem informação relativa ao cliente, ao tipo de mercadoria, ao serviço prestado e ao custo do mesmo. Os dados foram obtidos a partir da extração e processamento dos milhões de registos individuais, registados em guias de transporte emitidas pela empresa no período em questão.

A fase inicial passou pelo processamento dos registos em bruto, de forma a transformá-los em dados sobre os quais fosse possível aplicar ferramentas de análise, no sentido de deles extrair informação e conhecimento. Os dados extraídos do *software* de faturação vêm em formato *Microsoft Excel*, sendo ficheiros com vários milhares de linhas e mais de meia centena de colunas. Estes dados têm muitos registos, sendo uma grande parte destes importantes para o estudo da atividade da lavandaria, principalmente os relativos à data de emissão, à quantidade, ao cliente e ao total líquido. No entanto, os dados relativos ao tipo de peça e ao tipo de lavagem são registados num campo de texto livre. Este campo é editado manualmente pelo operador aquando da emissão de cada guia de transporte. De uma forma geral é registado o número da encomenda, o tipo de peça e o tipo de lavagem. Sendo um campo de registo manual, encontra-se sujeito a erro humano, o que dificultou de alguma forma o processamento dos dados.

Assim, para que fosse possível obter os dados, essenciais à realização deste estudo, relativos aos tipos de peça e de lavagem, foi necessário criar uma ferramenta de processamento dos dados registados num campo de texto livre, que iterativamente ia extraíndo as informações relativas às peças e lavagens. Conforme referido, esta tarefa foi prejudicada por erros nos registos de texto livre, que levaram a um grande consumo de tempo

na identificação e correção dos mesmos, de forma a permitir que a ferramenta de extração da informação relevante fosse eficaz na extração da informação pretendida. Este processo resultou na identificação das diferentes tipologias de lavagens e nos diferentes tipos de peças processados nesta lavandaria industrial, tornando possível a aplicação dos métodos de análise que são aplicados no decurso deste estudo.

Assim, depois de extraídas as guias de transporte, os dados contidos nas mesmas foram trabalhados e organizados em Excel de forma a obter-se uma tabela de dados com a seguinte configuração:

Data de emissão	Tipo de Lavagem	Tipo de Peça	Quantidade	Cliente	Total Líquido (€)
03/01/2018	ANTI-PILLING	FELPA	2	196	0,7
03/01/2018	TUMBLER	FELPA	263	20	15,78
⋮	⋮	⋮	⋮	⋮	⋮
29/12/2023	NORMAL	SWEAT	35	282	14,00

Tabela 2 - Tabela de dados após processamento

O campo **data de emissão** corresponde à data em que determinada guia de transporte foi lançada e emitida, na conclusão de determinado serviço.

O campo **tipo de lavagem** descreve o tipo de procedimento, ou serviço, que foi prestado, existindo os seguintes **20** tipos de lavagem diferentes:

Tipo de Lavagem			
60°	ANTI-PILLING C/SILICONE	FORTE	PROGRAMA 40
ANTI-CORROSÃO	C/DISPERSANTE	MICRONIZAR	SAND WASH
ANTI-ESTÁTICO	ENZIMÁTICA	NORMAL	SECO
ANTI-OLEO	ENZIMÁTICA C/SILICONE	NORMAL AMOSTRAS	STONE-WASH
ANTI-PILLING	ESPECIAL 60°	NORMAL C/SILICONE	TUMBLER

Tabela 3 - Tipos de lavagem

O campo **tipo de peça** indica qual a tipologia de peça têxtil, onde foi aplicado o tipo de lavagem, existindo **43** tipos diferentes:

Tipo de Peça			
AMOSTRAS	CORDÃO	MALHAS	SAIA
BLAZER	CORTINAS	MÁSCARAS	SINGLET
BLUSA	CREPOM	MEIAS	SUJOS
BODY	CUECA	NAO DEFINIDO	SWEAT
CALÇA	FELPA	PAINEIS	TIRAS
CALÇÃO	FITAS	PANOS	TOP
CAMISA	GOLAS	PIJAMAS	T-SHIRT
CAMISOLA	HOODIES	POLO	TUBO
CAPA	LEGGINS	PUFS	TUNICA
CASACO	LENÇOS	RENDAS	VESTIDO
COLETES	MACACÃO	ROBE	

Tabela 4 - Tipos de peça

O campo **quantidade** indica o número de peças relativo ao serviço.

O campo **cliente** indica o código numérico correspondente ao cliente do serviço em questão, existindo 95 clientes, com os seguintes códigos:

Código de Cliente									
9	92	149	175	204	218	241	251	261	274
22	94	155	176	205	222	242	252	262	277
32	110	159	178	206	225	243	253	263	279
34	112	163	182	208	229	244	254	264	281
55	119	164	184	209	230	245	255	265	282
59	123	167	185	211	232	246	256	266	
62	127	169	190	213	233	247	257	269	
63	129	171	193	214	236	248	258	271	
64	135	172	194	215	237	249	259	272	
75	141	174	196	217	240	250	260	273	

Tabela 5 - Códigos de cliente

O campo **total líquido** indica o valor a faturar, em euros, relativo ao serviço prestado.

Os dados, depois de organizados da forma referida na Tabela 1, foram trabalhados como se explica a seguir.

O objeto de investigação foi delimitado à análise do total líquido por mês e da sua variação ao longo do período em análise, considerando para tal a relação com os Clientes, os Tipos de Peça e os Tipos de Lavagem. Assim, foi necessário, a partir dos milhões de casos na Tabela 1, calcular o total líquido por mês para cada cliente, para cada tipo de peça e para cada tipo de lavagem. Como se pretendia uma análise relativa aos seis anos entre 2018 e 2023, foi necessário considerar 72 meses, P1, P2, P3, ..., P72.

Neste contexto, os dados a considerar passaram a ser os constantes nas três tabelas seguintes:

Cliente	P1	P2	P3	P4	P5	P6	...	P72
34	5256,65	4151,89	3839,76	5568,04	5211,26	7196,83	...	5995,42
176	924,21	764,94	1856,16	256,73	2723,27	3368,74	...	6680,08
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
282	0	0	0	0	0	0	...	642,60

Tabela 6 - Total líquido mensal por cliente (95 clientes x 72 meses)

Tipo Peça	P1	P2	P3	P4	P5	P6	...	P72
T-SHIRT	20235,86	18200,14	13850,69	11174,42	26996,26	18533,93	...	20564,00
POLO	5157,24	6619,73	4374,99	3202,06	2866,53	6493,14	...	4616,04
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
COLETES	0	0	0	0	0	0	...	0

Tabela 7 - Total líquido mensal por tipo de peça (43 peças x 72 meses)

Tipo Lavagem	P1	P2	P3	P4	P5	P6	...	P72
NORMAL	7133,33	4952,84	4315,61	5434,31	6913,53	7885,54	...	11714,92
ESPECIAL 60°	2923,26	3275,42	2515,01	2485,76	3649,04	3418,36	...	6785,19
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
FORTE	0	0	0	1,6	0	0	...	0

Tabela 8 - Total líquido mensal por tipo de lavagem (20 lavagens x 72 meses)

Em cada uma das tabelas, cada linha de valores corresponde a uma série temporal. A título de exemplo, a Figura 18 mostra as séries do total líquido para os cinco clientes mais importantes para a empresa em termos de faturação:

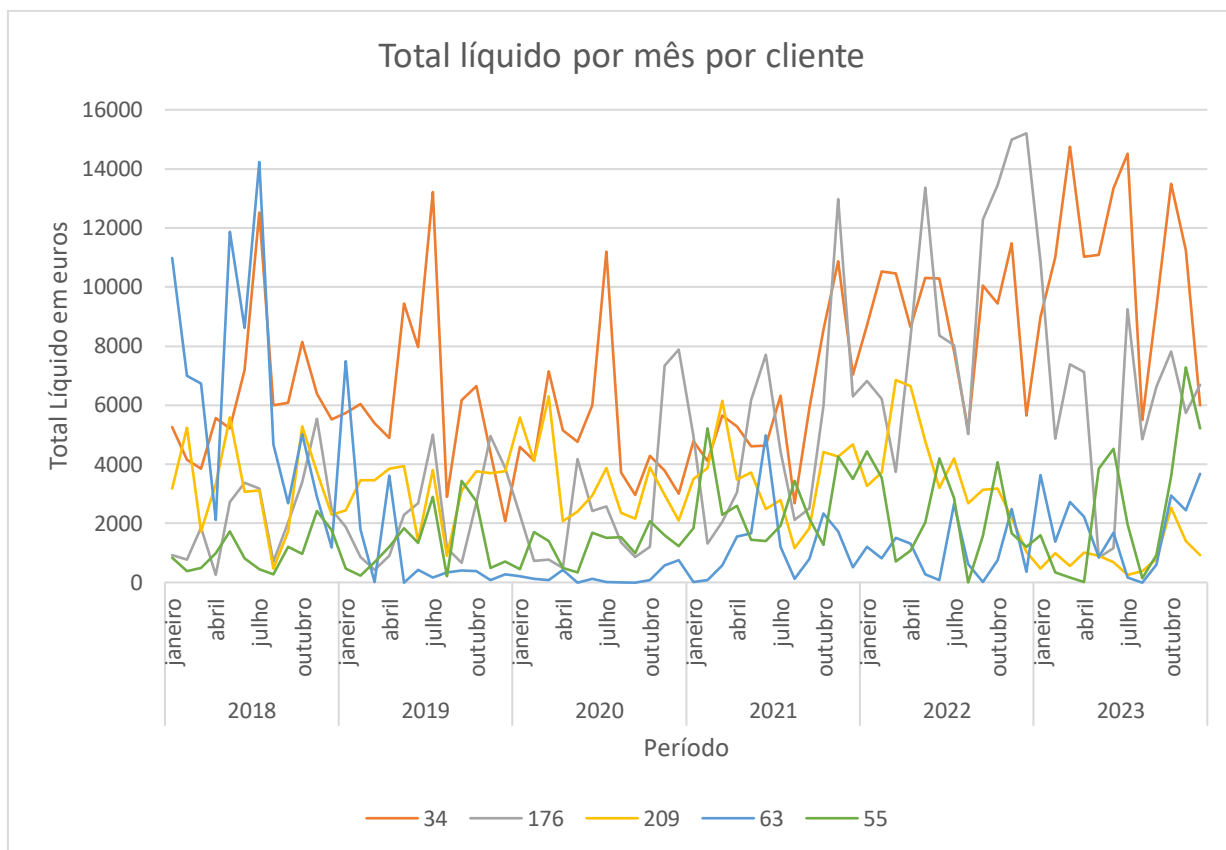


Figura 18 - Variação do total líquido para os cinco clientes mais importantes em termos de faturação

Os dados de 2018 a 2022 foram usados para *clustering* e ajuste de modelos de previsão e os de 2023 para teste destes últimos.

5.2. Clustering

Conforme referido na secção anterior, os dados de 2018 a 2022 foram considerados no sentido de encontrar similaridades entre diferentes tipos de clientes, lavagens e peças, agrupando-os em função do total líquido. Os *clusters* resultantes foram depois analisados.

A metodologia adotada passou pela aplicação de dois algoritmos de *clustering*, utilizando diferentes fórmulas de cálculo de distância. Assim, numa primeira fase, foi considerado o algoritmo k-médias, utilizando-se a distância Euclidiana como referência para a distância entre as séries temporais. Numa segunda fase, foi utilizado o algoritmo *k-medóides*, combinado com a distância obtida através do DTW. Nas duas fases, usou-se o coeficiente silhueta para determinar o número ótimo de *clusters*.

Atendendo ao facto de que os dados registados por vezes registam uma variação de escala significativa, tendo em conta que determinados períodos têm valores de total líquido muito superiores a outros, procedeu-se também à aplicação dos referidos algoritmos de *clustering*, aos dados standardizados, comparando os resultados obtidos com os dos dados não standardizados.

De forma resumida, foram realizadas as seguintes análises:

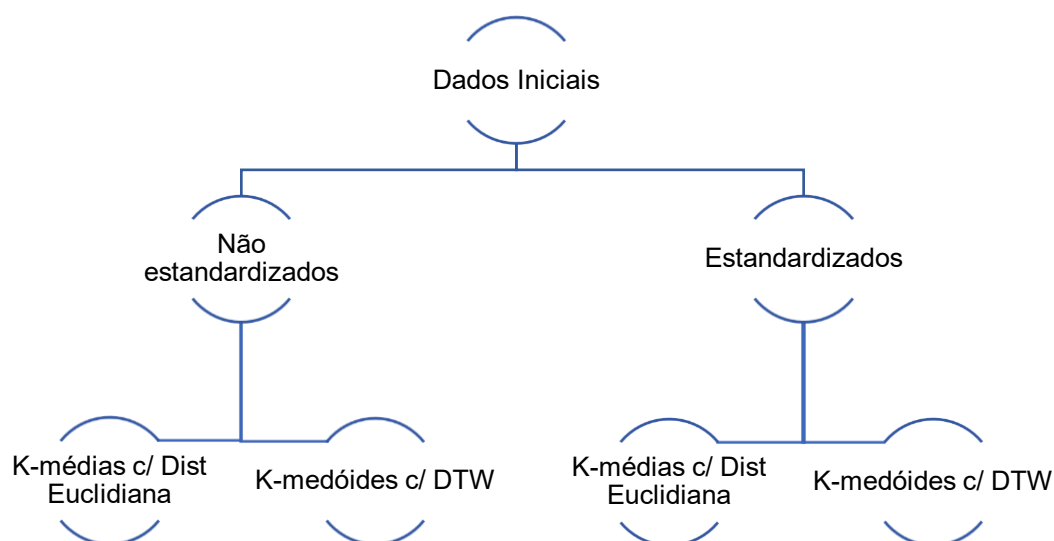


Figura 19 - Aplicação dos algoritmos de clustering aos dados não estandardizados e estandardizados

Esta metodologia foi aplicada aos três conjuntos de dados do total líquido dos clientes, tipos de peça e tipos de lavagem.

A estandardização e aplicação dos algoritmos de *clustering* foi feita com recurso ao *software* Matlab (Ghazal et al., 2021; Wegmann, Dominique, et al., 2021), com a *Statistics and Machine Learnign Toolbox*, tendo sido obtidos os resultados que se apresentam no capítulo seguinte.

5.3. Previsão de séries temporais

A metodologia seguida assentou na análise dos dados relativos ao total líquido, sendo que os valores mensais registados ao longo do período de 2018 a 2022 foram usados para definir ou ajustar os modelos de previsão e os dados relativos ao ano de 2023 utilizados para testar os mesmos.

Assim, numa primeira fase foram organizados os dados de forma a obter-se uma tabela com a distribuição do total líquido por mês (ver secção 5.1). Em seguida, a estes dados foram aplicados vários modelos de previsão, sendo no final os resultados obtidos comparados entre si, recorrendo às métricas MAE e MAPE introduzidas anteriormente.

Para esta análise foram definidos oito modelos, tendo os mesmos sido aplicados, num primeiro momento à previsão do total líquido para o ano 2023 de forma global, ou seja, sem qualquer agrupamento da informação por cliente, peça ou tipo de lavagem. Num momento posterior, os modelos de previsão foram aplicados aos *clusters* obtidos na análise

de *clustering* elaborada no capítulo 5. Esta análise teve em conta os *clusters* relativos ao total líquido por cliente, por peça e por tipo de lavagem.

Os modelos aplicados são apresentados seguidamente.

5.3.1. Métodos simples

Conforme apresentado em capítulo anterior, existem modelos que, apesar de simples, por vezes se constituem como os melhores para a previsão do comportamento de séries temporais. Assim, foram aplicados aos dados em análise os métodos Mean, Naive, SNaive e Drift. O *software* utilizado foi o Microsoft Excel.

5.3.2. Métodos regressivos e autoregressivo

A aplicação deste tipo de modelos mais complexos poderá permitir uma melhor adaptação aos dados disponíveis e com previsões mais certas.

Deste modo, foram considerados os quatro modelos regressivos e um autorregressivo descritos adiante.

Estes modelos tiveram em conta o total líquido em cada um dos doze meses do ano, bem como os períodos correspondentes ao registo da série temporal, ou seja, os 60 (5 anos * 12 meses) períodos correspondentes a cada um dos registos mensais, e ainda o ano a que corresponde cada leitura, de 1 a 5, correspondendo a 2018 a 2022, respetivamente.

Na elaboração deste estudo, foi utilizado o *software* Microsoft

, através da ferramenta Análise de Dados, que inclui uma opção específica para regressão linear. Este *software* utiliza o **método dos mínimos quadrados** (*ordinary least squares* - OLS) para estimar os coeficientes de regressão descritos. O *software* R foi usado para o modelo autorregressivo.

5.3.2.1. Modelo 1

De forma a integrar efeitos de sazonalidade e da possível tendência quadrática ao longo dos anos, este modelo 1 é representado pela seguinte equação de regressão:

$$\hat{y} = f(Mês_1, Mês_2, \dots, Mês_{11}, Ano, Ano^2)$$
$$\hat{y} = \beta_0 + \beta_1 \times Mês_1 + \beta_2 \times Mês_2 \dots + \beta_{11} \times Mês_{11} + \beta_{12} \times Ano + \beta_{13} \times Ano^2$$

onde $Mês_i = 1$ se o mês em consideração foi o i e $Mês_i = 0$ se o mês em consideração não for o i .

5.3.2.2. Modelo 2

De forma a integrar efeitos de sazonalidade e da possível tendência quadrática ao longo dos períodos, este modelo 2 é representado pela seguinte equação de regressão:

$$\hat{y} = f(Mês_1, Mês_2, \dots, Mês_{11}, Período, Período^2)$$
$$\hat{y} = \beta_0 + \beta_1 \times Mês_1 + \beta_2 \times Mês_2 \dots + \beta_{11} \times Mês_{11} + \beta_{12} \times Período + \beta_{13} \times Período^2$$

5.3.2.3. Modelo 3

De forma a integrar efeitos de sazonalidade e da possível tendência linear ao longo dos anos, este modelo 3 é representado pela seguinte equação de regressão:

$$\hat{y} = f(Mês_1, Mês_2, \dots, Mês_{11}, Ano)$$
$$\hat{y} = \beta_0 + \beta_1 \times Mês_1 + \beta_2 \times Mês_2 \dots + \beta_{11} \times Mês_{11} + \beta_{12} \times Ano$$

5.3.2.4. Modelo 4

De forma a integrar efeitos de sazonalidade e da possível tendência linear ao longo dos períodos, este modelo 4 é representado pela seguinte equação de regressão:

$$\hat{y} = f(Mês_1, Mês_2, \dots, Mês_{11}, Período)$$
$$\hat{y} = \beta_0 + \beta_1 \times Mês_1 + \beta_2 \times Mês_2 \dots + \beta_{11} \times Mês_{11} + \beta_{12} \times Período$$

5.3.2.5. Modelo ARIMA

O método ARIMA foi utilizado no sentido de procurar resultados mais precisos, recorrendo para tal a um método mais complexo, mas possivelmente mais eficaz. A descrição formal é a apresentada na secção 3.2.2.6. Foi considerado o caso mais geral de uma eventual sazonalidade.

5.3.3. Análise de resíduos

Uma vez ajustados os modelos aos dados, calculou-se os resíduos para cada um deles (ver o apêndice correspondente à secção 9.5). De acordo com (Hyndman &

Athanasopoulos, 2018), a média dos resíduos deve ser igual a zero, caso contrário as previsões são enviesadas. Além disso, os autores referem que é desejável, mas não necessário, que os resíduos tenham variância constante e que sejam normalmente distribuídos. Como não é necessário e havia limitações de tempo dado o prazo de entrega da dissertação, optou-se por não fazer uma análise residual aprofundada, que será deixada para trabalho futuro. Deste modo, testou-se apenas se a média dos resíduos é significativamente diferente de zero. Para tal, fixou-se um nível de significância igual a $\alpha=0,05$ e usou-se o teste t. A conclusão é que a média pode ser considerada significativamente diferente de zero se o seu valor absoluto for maior do que duas vezes o desvio-padrão a dividir pela raiz quadrada de n. A descrição do teste pode ser encontrada em (D. Anderson et al., 2010).

No caso dos modelos regressivos e autorregressivo, foram conduzidos testes de significância também para um nível de significância $\alpha=0,05$.

No caso dos modelos regressivos 1 a 4, usou-se o teste F para testar se existe ou não uma relação estatisticamente significativa entre a variável dependente e o conjunto de todas as variáveis independentes, ou seja, testou-se a significância global dos modelos. No caso de o teste F mostrar que o modelo testado é significativo, usou-se o teste t para determinar se cada uma das variáveis independentes no modelo é ou não significativa. Nos dois testes, a significância existe se $p<0,05$ e não existe caso contrário. Ambos estão descritos em (D. Anderson et al., 2010).

No caso do modelo ARIMA, usou-se o teste de Box-Ljung para testar se o modelo é ajustado ou não. O modelo é desajustado se $p<0,05$ e ajustado caso contrário. O teste está descrito em (Hyndman & Athanasopoulos, 2018).

Por motivos de limitação de tempo, não foi possível realizar uma análise detalhada dos resíduos dos modelos de previsão desenvolvidos. Contudo, com o objetivo de garantir a transparência do processo e permitir futuras avaliações, os resíduos foram devidamente registados e encontram-se disponíveis para consulta em apêndice, na secção 9.5. A análise preliminar que foi possível realizar é referida do no próximo capítulo.

6. Resultados e discussão

As diferentes ferramentas de análise de dados foram aplicadas ao conjunto de dados extraídos, passando numa primeira fase por uma análise primária à informação contida nos mesmos, de forma a permitir uma melhor perceção da operação da lavandaria industrial. De seguida, foram aplicados os algoritmos de *clustering* indicados no capítulo anterior, procedendo-se à análise detalhada de cada um dos *clusters* obtidos. Por fim, foi realizada a previsão de séries temporais, quer dos dados globais quer dos dados divididos pelos *clusters* anteriormente obtidos, tendo sido considerados os modelos referidos no capítulo precedente.

6.1. Análise primária de dados

Os dados relativos à faturação da empresa foram recolhidos relativamente aos anos 2018, 2019, 2020, 2021, 2022 e 2023, nomeadamente:

- Data de Emissão;
- Tipo de Lavagem – 20 tipos;
- Tipo de Peça – 43 tipos;
- Quantidade;
- Preço Unitário;
- Número do cliente – 93 clientes;
- Total Líquido;

Depois de organizados os dados, realizou-se uma análise primária, relativa ao funcionamento da empresa, da qual seguidamente são apresentados os resultados. Torna-se claro que o essencial do negócio está concentrado em poucos clientes, tipos de lavagem e de peça

6.1.1. Total líquido por cliente

O total líquido no período em análise foi de 2.216.232,78 €, dividindo-se da seguinte forma pelos anos:

- 2018 – 349.890,33 €;
- 2019 – 309.911,57 €;
- 2020 – 274.700,55 €;
- 2021 – 362.971,80 €;
- 2022 – 503.863,90 €;
- 2023 – 414.894,63 €.

A figura seguinte ilustra esta informação, sendo claro que o valor diminui até 2022, aumentou daí até 2022 e depois voltou a diminuir.

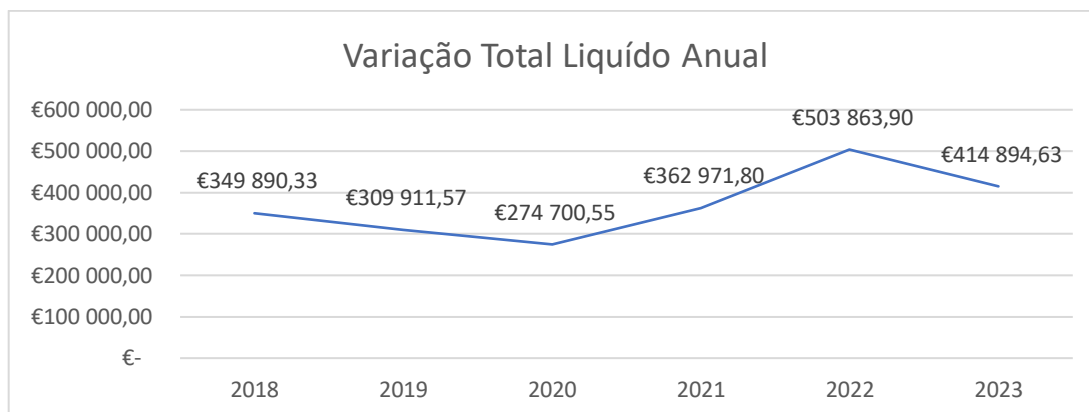


Figura 20 - Variação do total líquido anual

Em termos de distribuição do total líquido por cliente, verifica-se que, no período em análise, 90 % deste valor se encontra distribuído por 15 clientes (16%) e que 54,9% se encontra distribuído por 4 clientes (4%).

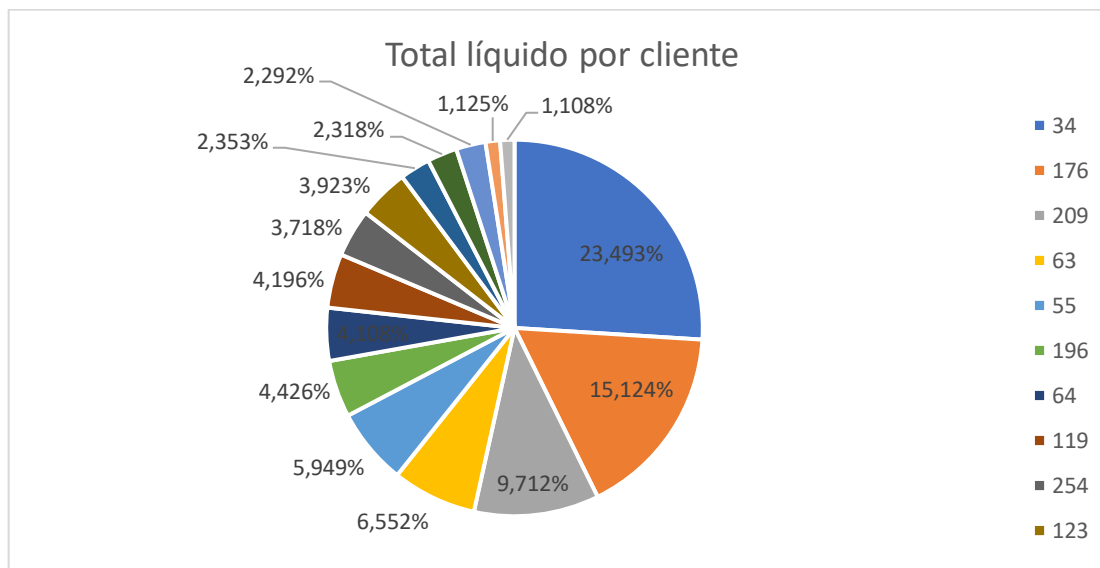


Figura 21 - Distribuição do total líquido por cliente (correspondentes a 90 % do total líquido)

A variação do total líquido por cliente tem diferentes comportamentos ao longo do período em análise, para cada um dos clientes. De seguida apresenta-se a representação gráfica desta variação para os quatro clientes mais relevantes.

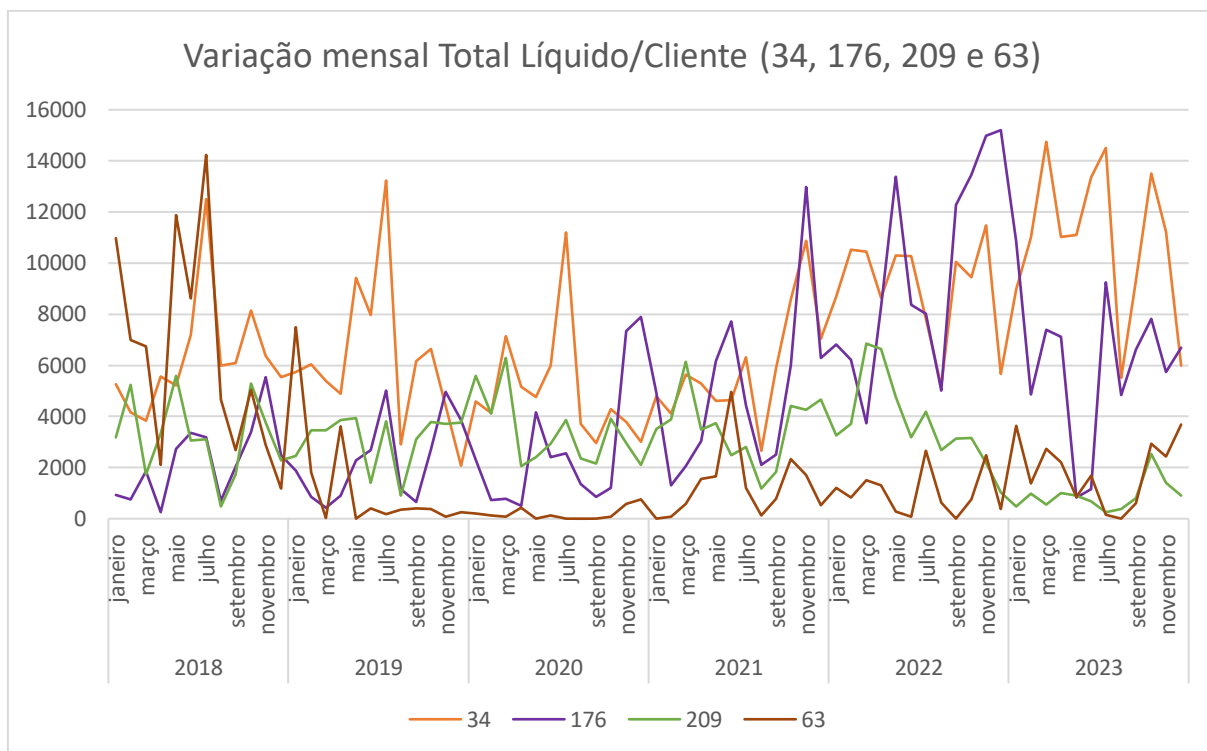


Figura 22 - Variação do total líquido por cliente (correspondente a 54,9 % do total líquido)

Da análise da Figura 22 extrai-se que cliente 34 apresenta o maior valor de total líquido, com sazonalidade, com picos elevados e quedas bruscas, especialmente entre 2018 e 2019, verificando-se crescimento a partir de 2021. O cliente 176 demonstra crescimento mais consistente a partir de 2020 até 2022, apresentando também sazonalidade, com diminuição do valor em 2023. Os clientes 209 e 63 mantêm-se mais estáveis ao longo do período, com valores geralmente mais baixos e menos oscilações, sendo o cliente 63 o menos expressivo.

6.1.2. Total líquido por tipo de lavagem

Dos 20 tipos de lavagem existentes, 92,6 % do total líquido corresponde a 6 tipos de lavagem (30%). Os dois tipos de lavagem NORMAL e a ESPECIAL 60° (10%) correspondem a 57,0 % do total líquido no período.

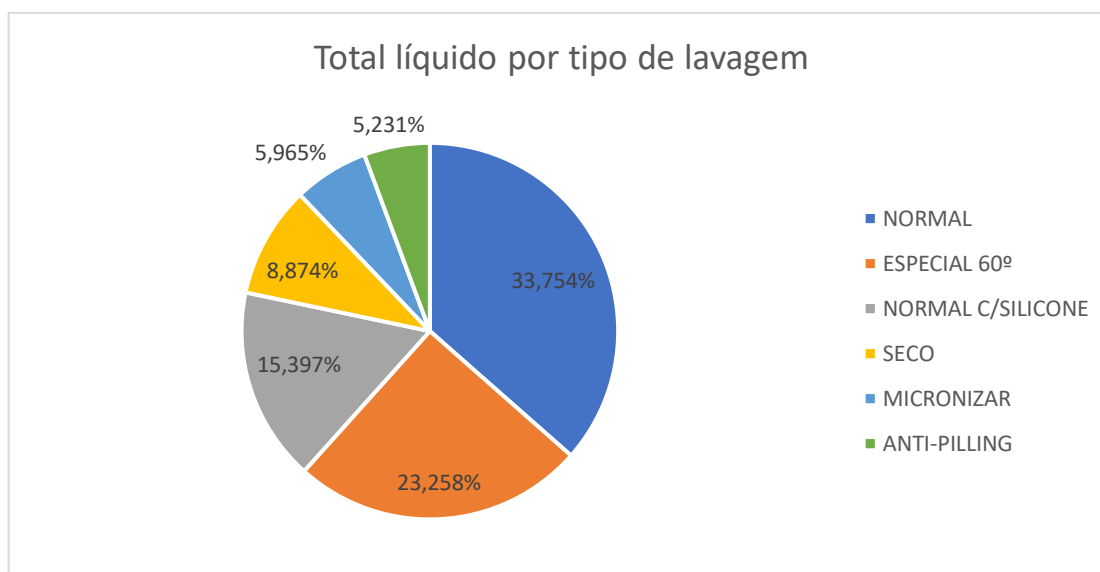


Figura 23 - Distribuição do total líquido por tipo de lavagem (correspondente a 92,6 % do total líquido)

A variação do total líquido por tipo de lavagem tem diferentes comportamentos ao longo do período em análise, para cada um dos tipos de lavagem. De seguida apresenta-se a representação gráfica desta variação para as seis lavagens mais importantes.

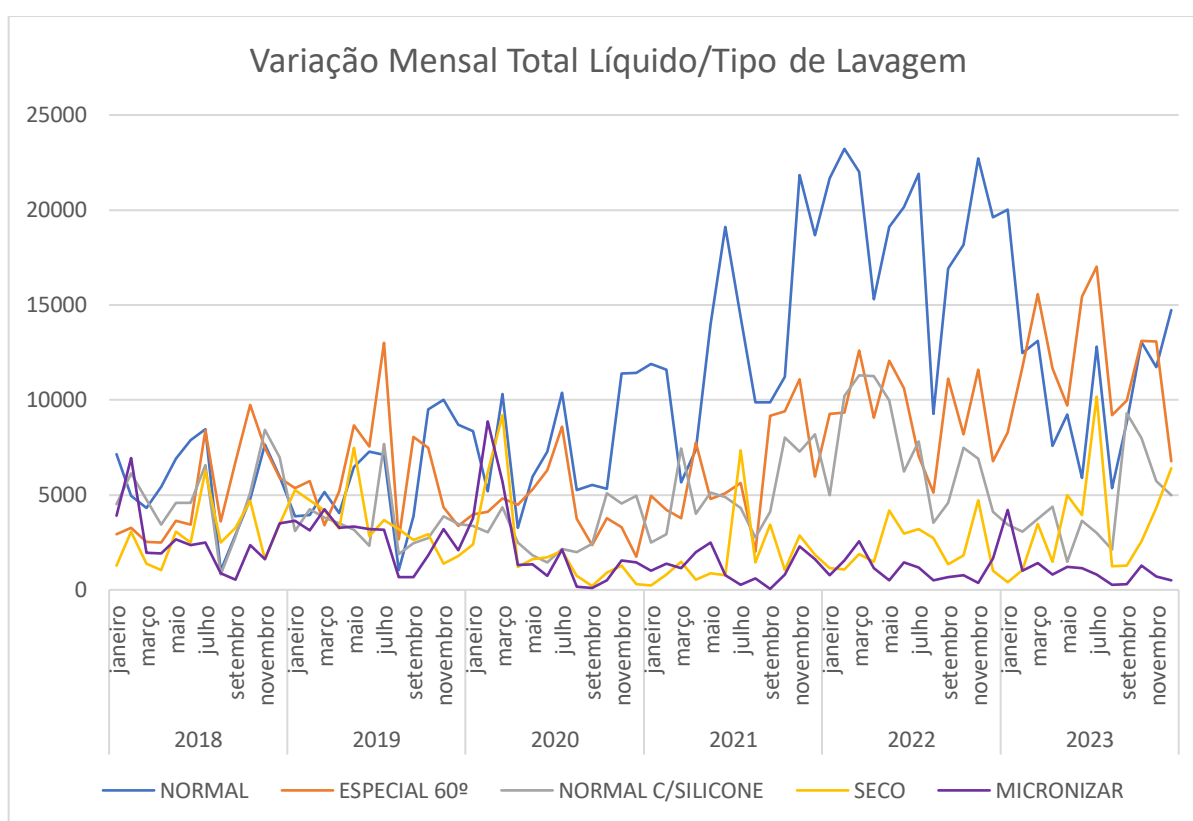


Figura 24 - Variação do total líquido por tipo de lavagem (correspondente a 92,6 % do total líquido)

A variação do tipo de lavagem NORMAL entre 2018 e 2023, apresenta uma tendência geral de crescimento até meados de 2021, seguida de uma forte volatilidade. Entre 2018 e 2019, os valores são relativamente baixos e estáveis, situando-se em torno de 5000 a 10000

euros. A partir de 2020, observa-se um crescimento mais acentuado, atingindo picos próximos a 25000€ no final de 2021. Contudo, a partir de 2022, há uma queda significativa com oscilações marcantes e valores reduzidos, embora em 2023 se note uma leve recuperação. O tipo de lavagem ESPECIAL 60° apresenta um padrão de crescimento gradual com oscilações significativas. Entre 2018 e 2019, os valores variam entre 2000 e 10000 euros, com períodos de quedas seguidos por recuperações rápidas. A partir de 2020, há uma tendência de aumento mais consistente, com picos acima de 15000€ entre 2021 e 2023. Contudo, a série mantém um comportamento irregular, alternando períodos de forte subida e descida. O tipo NORMAL C/SILICONE apresenta um comportamento estável entre os anos 2018 e 2021, com variações em diferentes meses entre os 1000€ e os 8000€, em 2022 existe um grande pico seguido de um decréscimo até ao quatro trimestre de 2023, onde volta a subir. O tipo de lavagem SECO apresenta um desempenho constante, com picos ao longo do período em análise, mas com uma tendência de estabilidade. O tipo MICRONIZAR tem tendência decrescente entre 2018 e 2023. No entanto, nos anos 2018 e 2022 teve picos muito elevados, na ordem dos 8000€, mantendo estabilidade nos restantes momentos.

6.1.3. Total líquido por tipo de peça

Dos 43 tipos de peças existentes, 91,8 % do total líquido corresponde a 8 tipos de peças (19%). Os tipos de peça T-SHIRT e POLO correspondem a 65,0 % do total líquido no período.

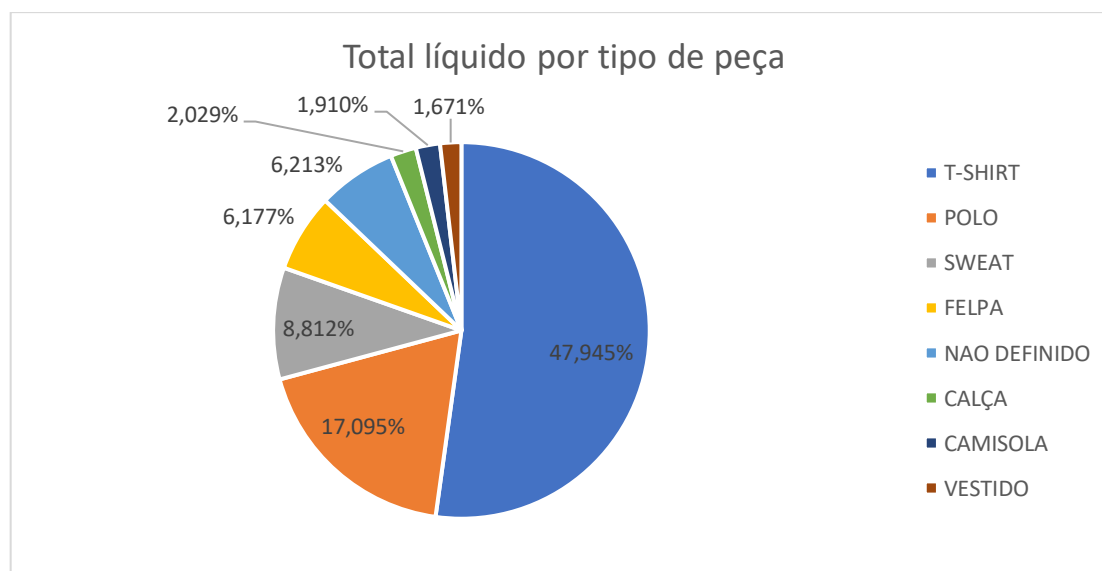


Figura 25 - Distribuição do total líquido por tipo de peça (correspondente a 91,8 % do total líquido)

A variação do total líquido por tipo de peças, apresenta comportamentos diferentes para cada um dos tipos de peça. O seguinte gráfico ilustra esta variação no período em análise.

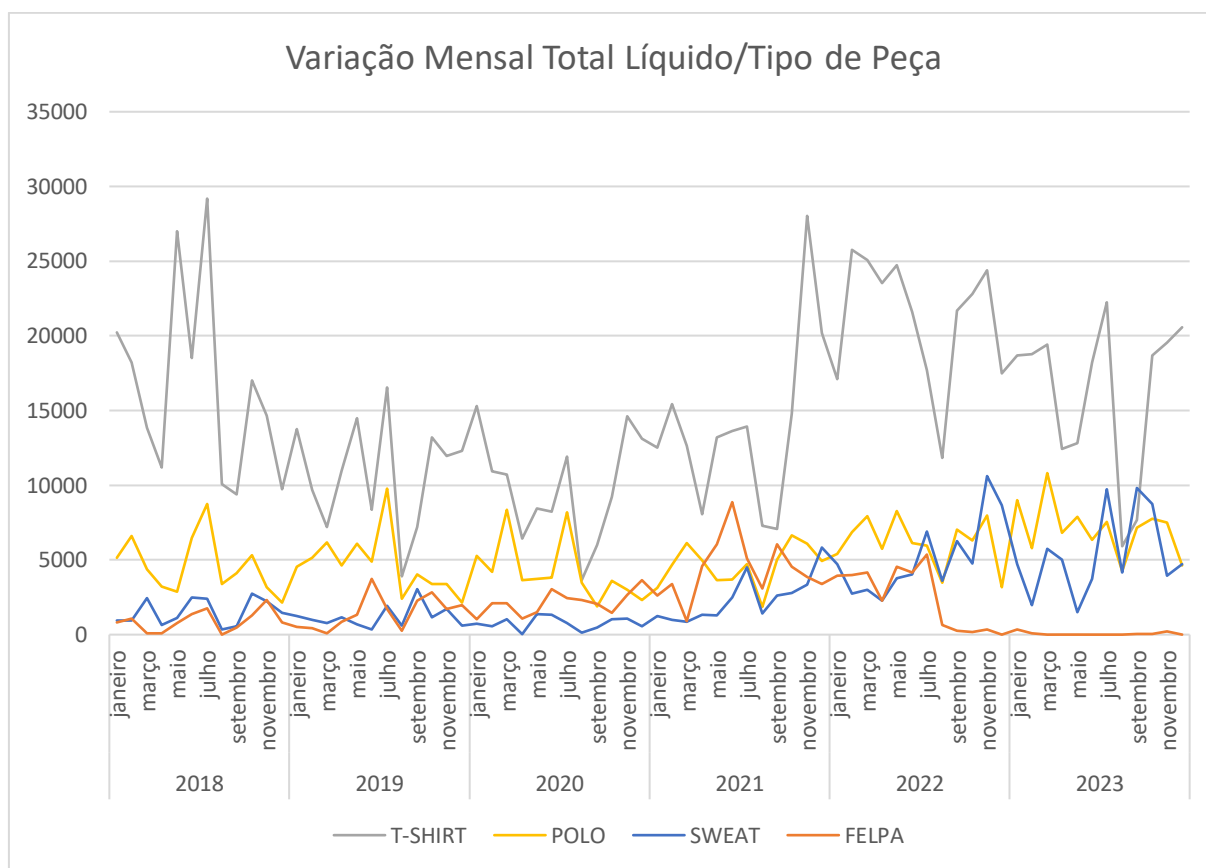


Figura 26 - Variação do total líquido por tipo de peça (correspondente a 79,3% do total líquido)

O total líquido relativo ao tipo de peça T-SHIRT manteve estabilidade, com uma queda entre os anos 2019 e 2021, seguindo-se uma subida acentuada no final de 2021, mantendo depois estabilidade com tendência decrescente até 2023. O tipo de peça POLO manteve estabilidade ao longo de todo o período, com tendência crescente, especialmente a partir de 2022. O tipo de peça SWEAT apresenta um desempenho regular até meados de 2021, data a partir da qual começa a apresentar um crescimento rápido, com alguns picos. O tipo de peça FELPA apresenta um desempenho constante, com algum crescimento até 2021, onde tem um pico de crescimento muito elevado. Após esse período, tem um desempenho constante até ao terceiro trimestre de 2022 onde começa a descer de forma acentuada, tornando-se quase residual.

6.2. Clustering

Na secção anterior foi realizada uma análise primária dos dados, obtendo-se algumas informações importantes sobre o funcionamento da lavandaria industrial. No entanto, para permitir a segmentação da informação, agrupando-a de acordo com as similaridades que tem, é necessário aplicar os algoritmos de *clustering*, de forma a obter os *clusters* relativos aos clientes, tipos de lavagem e tipos de peça.

6.2.1. Aplicação dos algoritmos

Nesta secção são mostrados os resultados da aplicação dos algoritmos apresentados na secção 3.1 aos dados processados conforme referido na secção 5.1, no sentido de se obterem informações sobre o número ótimo de *clusters* e a composição dos mesmos.

6.2.1.1. Algoritmo k-médias sem estandardização

O *clustering* foi feito com o algoritmo k-médias sem estandardização dos dados. O maior coeficiente silhueta está realçado em cada caso e permite identificar o número ótimo de clusters.

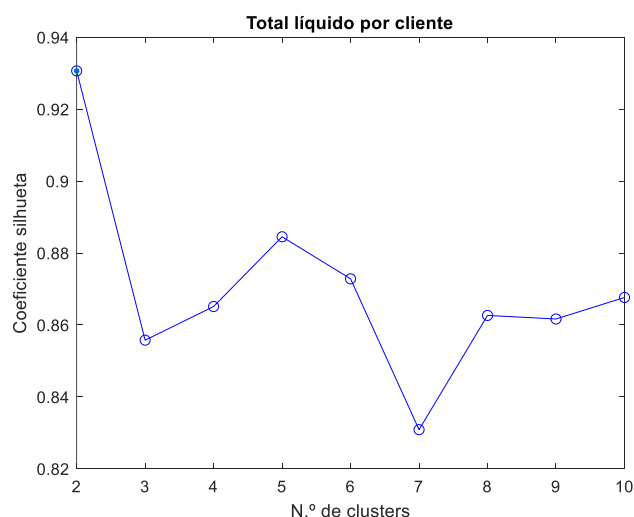


Figura 27 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias sem estandardização, total líquido por cliente.

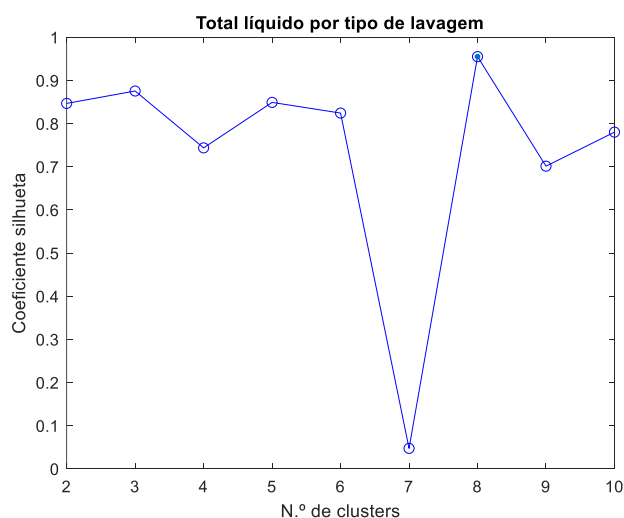


Figura 28 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias sem estandardização, total líquido por tipo de lavagem.

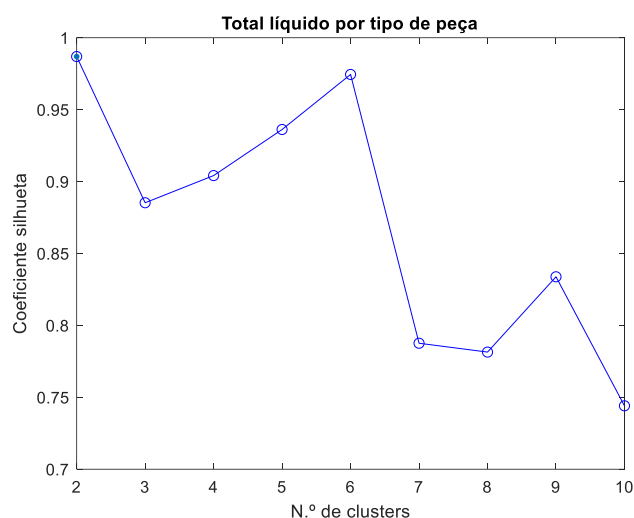


Figura 29 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias sem estandardização, total líquido por tipo de peça.

		Coeficiente Silhueta								
k		2	3	4	5	6	7	8	9	10
Total líquido	Cliente	0.9307	0.8558	0.8651	0.8845	0.8728	0.8309	0.8627	0.8617	0.8677
	Tipo de lavagem	0.8469	0.8757	0.7439	0.8493	0.8245	0.0474	0.9553	0.7015	0.7803
	Tipo de peça	0.9869	0.8853	0.9042	0.9362	0.9745	0.7876	0.7814	0.8338	0.7441

Tabela 9 – Coeficiente silhueta para diferentes valores do número k de clusters de clientes, tipos de lavagem e tipos de peça.

6.2.1.2. Algoritmo k-médias com estandardização

O *clustering* foi feito com o algoritmo k-médias com estandardização dos dados. O maior coeficiente silhueta está realçado em cada caso e permite identificar o número ótimo de *clusters*.

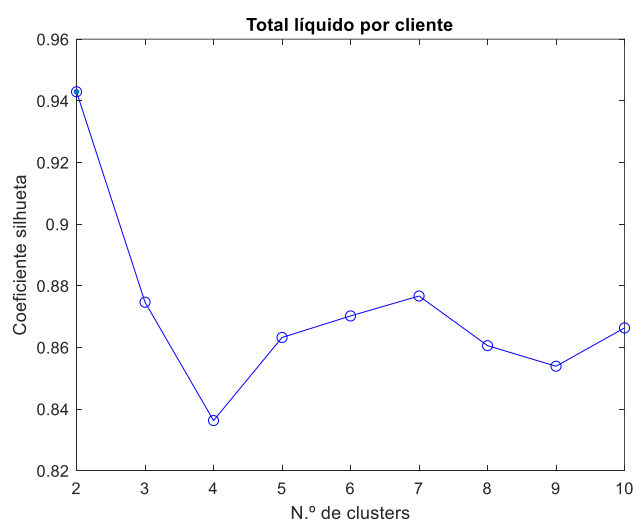


Figura 30 - Variação do coeficiente silhueta na aplicação do algoritmo k-médias com estandardização, total líquido por cliente.

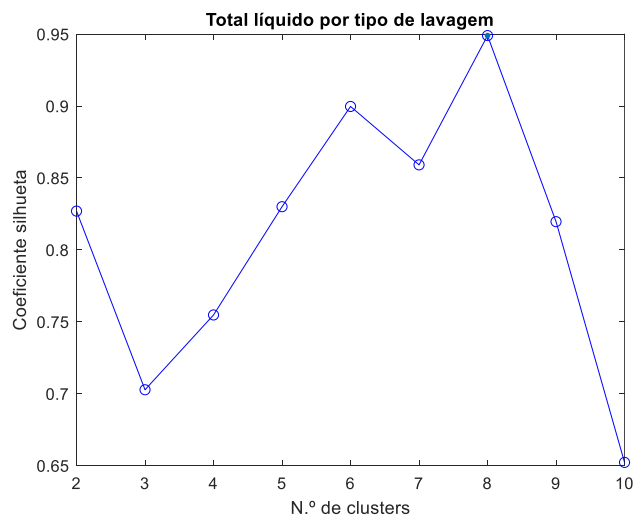


Figura 31 - Variação do coeficiente silhueta na aplicação do algoritmo *k*-médias com estandardização, total líquido por tipo de lavagem.

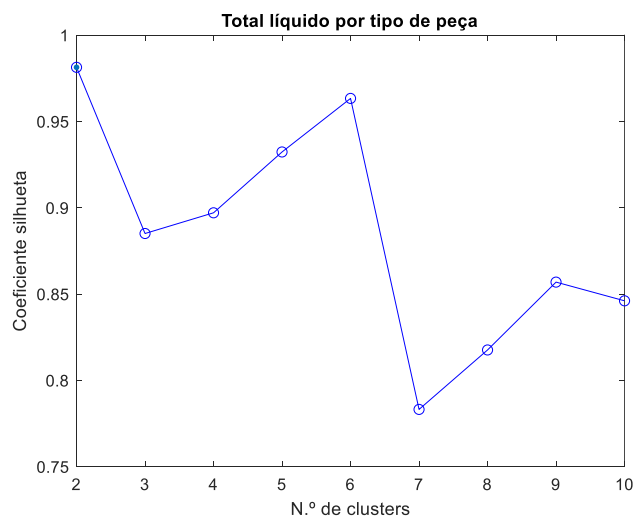


Figura 32 - Variação do coeficiente silhueta na aplicação do algoritmo *k*-médias com estandardização, total líquido por tipo de peça.

		Coeficiente Silhueta								
k		2	3	4	5	6	7	8	9	10
Total liquido	Cliente	0.9429	0.8747	0.8363	0.8632	0.8702	0.8767	0.8606	0.8539	0.8663
	Tipo de lavagem	0.8270	0.7028	0.7548	0.8300	0.8997	0.8591	0.9490	0.8196	0.6523
	Tipo de peça	0.9814	0.8851	0.8971	0.9323	0.9633	0.7832	0.8176	0.8569	0.8461

Tabela 10 - Coeficiente silhueta para diferentes valores do número *k* de clusters de clientes, tipos de lavagem e tipos de peça.

6.2.1.3. Algoritmo k-medóides sem estandardização

O *clustering* foi feito com o algoritmo k-medóides sem estandardização dos dados. O maior coeficiente silhueta está realçado em cada caso e permite identificar o número ótimo de *clusters*.

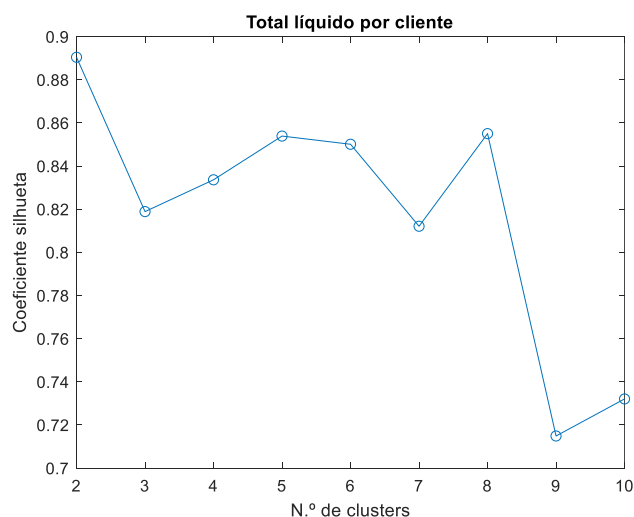


Figura 33 - Variação do coeficiente silhueta na aplicação do algoritmo k-medóides sem estandardização, total líquido por cliente.

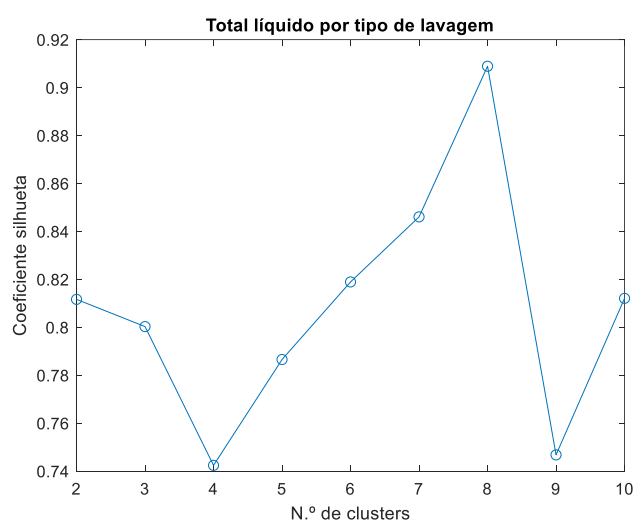


Figura 34 - Variação do coeficiente silhueta na aplicação do algoritmo k-medóides sem estandardização, total líquido por tipo de lavagem.

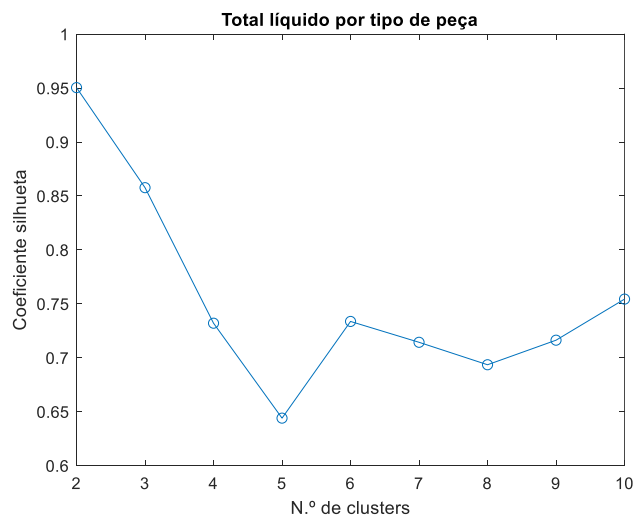


Figura 35 - Variação do coeficiente silhueta na aplicação do algoritmo *k-medóides* sem estandardização, total líquido por tipo de peça.

		Coeficiente Silhueta									
k		2	3	4	5	6	7	8	9	10	
Total líquido	Cliente	0,8904	0,8189	0,8336	0,8539	0,8501	0,8121	0,8550	0,7149	0,7321	
	Tipo de lavagem	0,8117	0,8003	0,7424	0,7866	0,8189	0,8461	0,9089	0,7468	0,8121	
	Tipo de peça	0,9504	0,8577	0,7320	0,6440	0,7337	0,7143	0,6935	0,7163	0,7544	

Tabela 11 - Coeficiente silhueta para diferentes valores do número *k* de clusters de clientes, tipos de lavagem e tipos de peça.

6.2.1.4. Algoritmo *k-medóides* com estandardização

O *clustering* foi feito com o algoritmo *k-medóides* com estandardização dos dados. O maior coeficiente silhueta está realçado em cada caso e permite identificar o número ótimo de *clusters*.

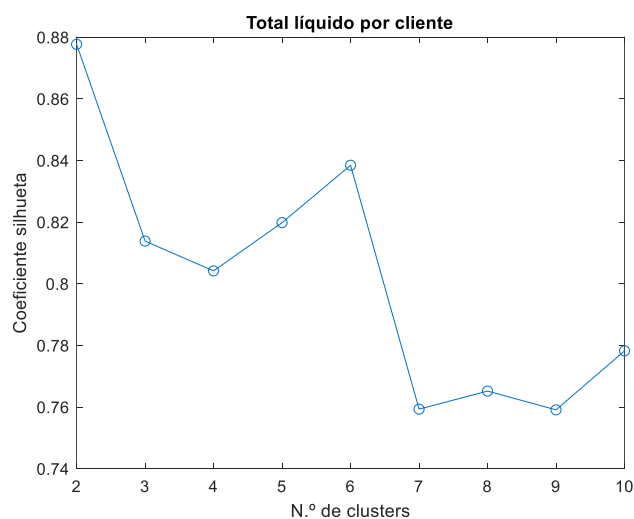


Figura 36 - Variação do coeficiente silhueta na aplicação do algoritmo *k-medóides* com estandardização, total líquido por cliente.

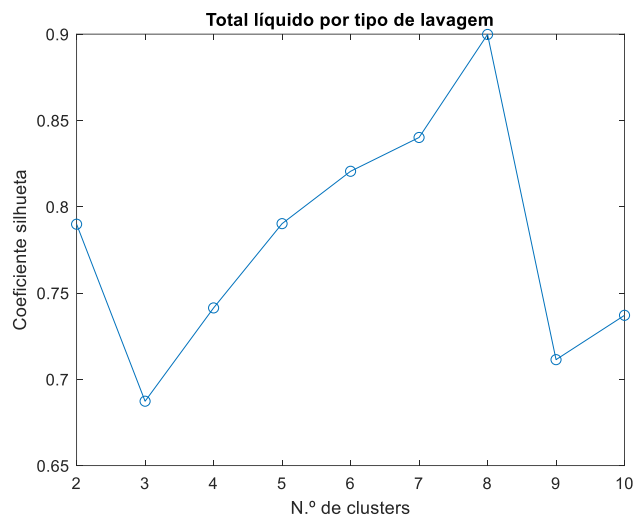


Figura 37 - Variação do coeficiente silhueta na aplicação do algoritmo *k-medóides* com estandardização, total líquido por tipo de lavagem.

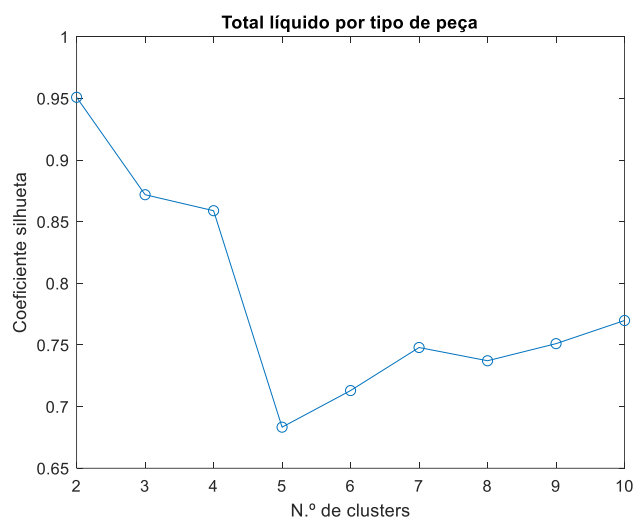


Figura 38 - Variação do coeficiente silhueta na aplicação do algoritmo *k-medóides* com estandardização, total líquido por tipo de peça.

		Coeficiente Silhueta								
k		2	3	4	5	6	7	8	9	10
Total líquido	Cliente	0,8777	0,8138	0,8042	0,8198	0,8384	0,7593	0,7652	0,7591	0,7782
	Tipo de lavagem	0,7899	0,6873	0,7414	0,7902	0,8205	0,8401	0,8998	0,7114	0,7371
	Tipo de peça	0,9509	0,8719	0,8589	0,6833	0,7130	0,7479	0,7372	0,7511	0,7698

Tabela 12 - Coeficiente silhueta para diferentes valores do número *k* de clusters de clientes, tipos de lavagem e tipos de peça.

6.2.2. Resumo dos resultados ótimos de *clustering*

O *clustering* dos dados foi efetuado com recurso a dois algoritmos diferentes, com e sem estandardização de dados e com a utilização de métodos de cálculo de distância

diferentes. Os diferentes métodos produziram resultados próximos, sendo, no entanto, diferentes. De seguida, apresenta-se um quadro resumo dos resultados ótimos obtidos no *clustering*, onde o número ótimo de *clusters* foi determinado considerando o número de *clusters* com maior coeficiente silhueta associado:

Algoritmo de <i>clustering</i> e medida de distância	Estandar dização	Total líquido por cliente		Total líquido por tipo de lavagem		Total líquido por tipo de peça	
		N.º ótimo de <i>clusters</i>	Coeficiente silhueta	N.º ótimo de <i>clusters</i>	Coeficiente silhueta	N.º ótimo de <i>clusters</i>	Coeficiente silhueta
K-médias com distância Euclidiana	Não	2	0,9307	8	0,9553	2	0,9869
	Sim	2	0,9429	8	0,9490	2	0,9814
K-medóides com dynamic time warping	Não	2	0,8904	8	0,9089	2	0,9504
	Sim	2	0,8777	8	0,8998	2	0,9509

Tabela 13 - Quadro resumo dos resultados obtidos no *clustering*.

Estes resultados são pormenorizados a seguir e mais adiante, na secção 6.2.3 é feita uma análise detalhada dos mesmos.

6.2.2.1. Total líquido por cliente

Os resultados obtidos por aplicação dos algoritmos de *clustering* indicam que o número ótimo de *clusters*, independentemente do algoritmo utilizado, é dois. Os *clusters* encontram-se bem separados, com o coeficiente silhueta a variar entre 0,8777 e 0,9429. O coeficiente silhueta mais elevado é obtido por aplicação do algoritmo k-médias, com distância Euclidiana e com estandardização de dados. Assim, esta é a abordagem mais adequada para agrupar clientes por total líquido. Neste caso, o *cluster* 1 compreende 92 casos, representando 96,84 % do total de 95 clientes, sendo que o *cluster* 2, compreende os restantes 3 casos, correspondendo a 3,16 % do total de clientes.

Algoritmo de <i>clustering</i> e medida de distância	Estandar dização	Total líquido por cliente		Tabela de frequências			
		N.º ótimo de <i>clusters</i>	Coeficiente silhueta	Cluster 1		Cluster 2	
				Casos	%	Casos	%
K-médias com distância Euclidiana	Não	2	0,9307	91	95,79%	4	4,21%
	Sim	2	0,9429	92	96,84%	3	3,16%
K-medóides com dynamic time warping	Não	2	0,8904	91	95,79%	4	4,21%
	Sim	2	0,8777	92	96,84%	3	3,16%

Tabela 14 - Quadro resumo do *clustering* dos clientes por total líquido.

6.2.2.2. Total líquido por tipo de lavagem

Os resultados obtidos por aplicação dos algoritmos de *clustering* indicam que o número ótimo de *clusters*, independentemente do algoritmo utilizado, é oito. Os *clusters* encontram-se bem separados, com o coeficiente silhueta a variar entre 0,8998 e 0,9553. O coeficiente silhueta mais elevado é obtido por aplicação do algoritmo k-médias, com distância Euclidiana e sem estandardização de dados. Deste modo, esta é a abordagem mais

adequada para agrupar os tipos de lavagem por total líquido. Neste caso, o *cluster* 1 compreende 13 casos, representando 65 % dos 20 tipos de lavagem, e todos os restantes sete *clusters* compreendem 1 caso cada, correspondendo a 5 % do total de tipos de lavagem.

Algoritmo de <i>clustering</i> e medida de distância	Estandarização	Total líquido por tipo de lavagem		Tabela de frequências			
		N.º ótimo de <i>clusters</i>	Coeficiente silhueta	<i>Cluster</i> 1		<i>Cluster</i> 2 a 8	
				Casos	%	Casos	%
K-médias com distância Euclidiana	Não	8	0,9553	13	65%	1	5%
	Sim	8	0,9490	13	65%	1	5%
K-medóides com dynamic time warping	Não	8	0,9089	13	65%	1	5%
	Sim	8	0,8998	13	65%	1	5%

Tabela 15 - Quadro resumo do *clustering* dos tipos de lavagem por total líquido.

6.2.2.3. Total líquido por tipo de peça

Os resultados obtidos por aplicação dos algoritmos de *clustering* indicam que o número ótimo de *clusters*, independentemente do algoritmo utilizado, é dois. Os *clusters* encontram-se bem separados, com o coeficiente silhueta a variar entre 0,9504 e 0,9869. O coeficiente silhueta mais elevado é obtido por aplicação do algoritmo k-médias, com distância Euclidiana e sem estandarização de dados. Assim, esta é a abordagem mais adequada para agrupar os tipos de peça por total líquido. Neste caso, o *cluster* 1 compreende 42 casos, representando 97,67 % do total de 43 tipos de peças, sendo que o *cluster* 2 compreende 1 caso, correspondendo a 2,33 % do total.

Algoritmo de <i>clustering</i> e medida de distância	Estandarização	Total líquido por tipo de peça		Tabela de frequências			
		N.º ótimo de <i>clusters</i>	Coeficiente silhueta	<i>Cluster</i> 1		<i>Cluster</i> 2	
				Casos	%	Casos	%
K-médias com distância Euclidiana	Não	2	0,9869	42	97,67%	1	2,33%
	Sim	2	0,9814	42	97,67%	1	2,33%
K-medóides com dynamic time warping	Não	2	0,9504	42	97,67%	1	2,33%
	Sim	2	0,9509	42	97,67%	1	2,33%

Tabela 16 - Quadro resumo do *clustering* dos tipos de peça por total líquido.

Após análise dos resultados obtidos por cliente, tipo de lavagem e peça, pode destacar-se que o algoritmo k-médias com distância Euclidiana, tanto com como sem estandarização, consistentemente apresenta coeficientes silhueta mais elevados do que o k-medóides com DTW. Isto sugere que o k-médias com distância Euclidiana é mais eficiente na separação dos *clusters* nos dados analisados.

A estandarização dos dados teve um impacto variado no coeficiente silhueta. Em alguns casos a estandarização melhorou o coeficiente silhueta, como nos clientes agrupados pelo total líquido com o algoritmo k-médias com distância Euclidiana. No entanto, noutros casos, a não estandarização apresentou melhores resultados, como nos tipos de lavagem e peça agrupados pelo total líquido com distância Euclidiana.

Em termos da consistência do número ótimo de *clusters*, destaca-se que para todos os métodos e métricas, o número ótimo de *clusters* foi consistente, o que sugere estabilidade nos dados e nos métodos utilizados.

6.2.3. Análise dos *clusters*

A secção anterior permitiu, através da aplicação de algoritmos de *clustering*, a definição do número ótimo de *clusters* para cada uma das situações em análise, relativas ao total líquido. No entanto, o *clustering* por si só não é suficiente para extrair informações relevantes. A análise dos *clusters* é importante de forma a permitir a caracterização dos mesmos e essa análise é feita em seguida.

6.2.3.1. Total líquido por cliente

Neste caso, conforme já foi referido antes, o algoritmo que apresentou melhores resultados foi o k-médias, com distância Euclidiana e os dados standardizados. Os resultados obtidos revelam que o número ótimo de *clusters* é dois, sendo que o *cluster* 1 compreende (96,84 %) e o *cluster* 2 compreende 3 casos (3,16 %).

A distribuição de cada um dos casos por *cluster* é a seguinte:

Código de Cliente									
9	92	149	175	204	218	241	251	261	274
22	94	155	176	205	222	242	252	262	277
32	110	159	178	206	225	243	253	263	279
34	112	163	182	208	229	244	254	264	281
55	119	164	184	209	230	245	255	265	282
59	123	167	185	211	232	246	256	266	
62	127	169	190	213	233	247	257	269	
63	129	171	193	214	236	248	258	271	
64	135	172	194	215	237	249	259	272	
75	141	174	196	217	240	250	260	273	

Tabela 17 - Distribuição dos clientes por *cluster*. A vermelho o *cluster* 1 e a azul o *cluster* 2.

Em termos gráficos, seguidamente é apresentado o comportamento da variação do total líquido por cliente, ao longo do período de cinco anos, ou seja, 60 meses, representando-se os diferentes *clusters* a cores diferentes.

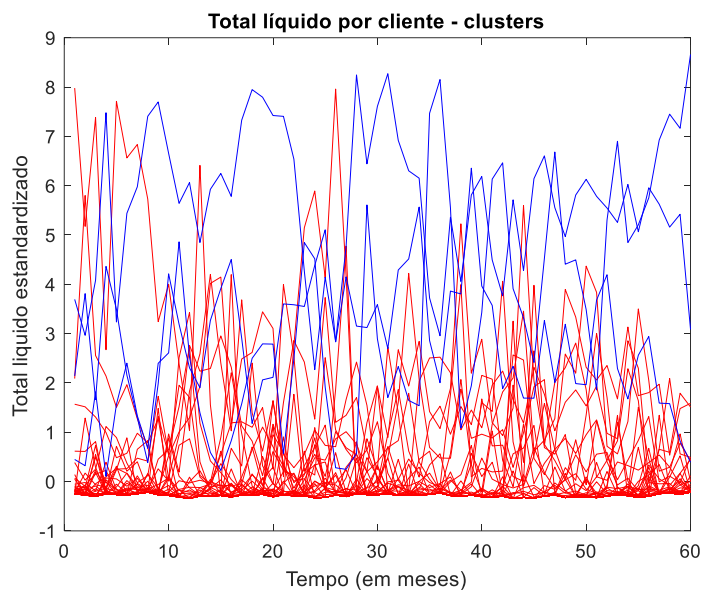


Figura 39 - Variação do total líquido por cliente. Cluster 1 a vermelho, cluster 2 a azul.

De seguida são representados graficamente os centróides dos dois *clusters*, ou seja, a média dos casos em cada um dos dois grupos. O centróide do *cluster* 1 está colorido a vermelho e o do *cluster* 2 a azul. É possível observar que o total líquido dos clientes do *cluster* 2 é, em média, consideravelmente maior do que o total líquido dos clientes do *cluster* 1. Portanto o enfoque da análise será no *cluster* 2.

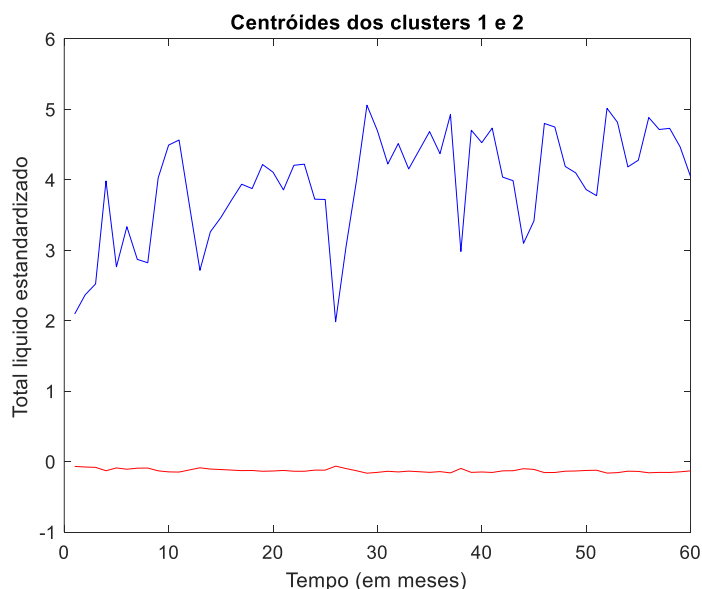


Figura 40 - Centróides dos clusters1 (vermelho) e 2 (azul).

Os clientes que englobam o *cluster* 2, representando 3,16 % dos clientes, são o 34, 176 e 209. A seguinte tabela apresenta os valores do total líquido de cada um destes clientes ao longo do período em análise.

Cliente	Total líquido (euros)						% em relação ao total de todos os clientes	% acumulada em relação ao total de todos os clientes
	2018	2019	2020	2021	2022	2018-2022		
34	75860	74874	60750	70428	108509	390421	21,674%	21,674%
176	27215	27417	32087	59545	115746	262011	14,545%	36,219%
209	38734	37650	40756	42387	44832	204360	11,345%	47,564%

Tabela 18 - Total líquido dos clientes do cluster 2 entre 2018 e 2022.

Conforme se pode verificar, apesar do número de clientes representar menos de 4 % do total de clientes da empresa, o valor do total líquido dos mesmos corresponde a mais de 47 % do total. O Cliente 34 é o que gera o maior volume de receitas, totalizando 390.421 euros entre 2018 e 2022, representando 21,674% do total de todos os clientes. O Cliente 176 contribui com 262.011 euros no mesmo período, o que corresponde a 14,545% do total. O Cliente 209 tem um total de 204.360 euros, representando 11,345% do total. Juntos, estes três clientes representam 47,564% do total de receitas no período de 2018 a 2022.

Os dados da tabela anterior permitem verificar que o comportamento de cada um dos clientes não é uniforme ao longo do período em análise e esta conclusão é reforçada pela figura seguinte:

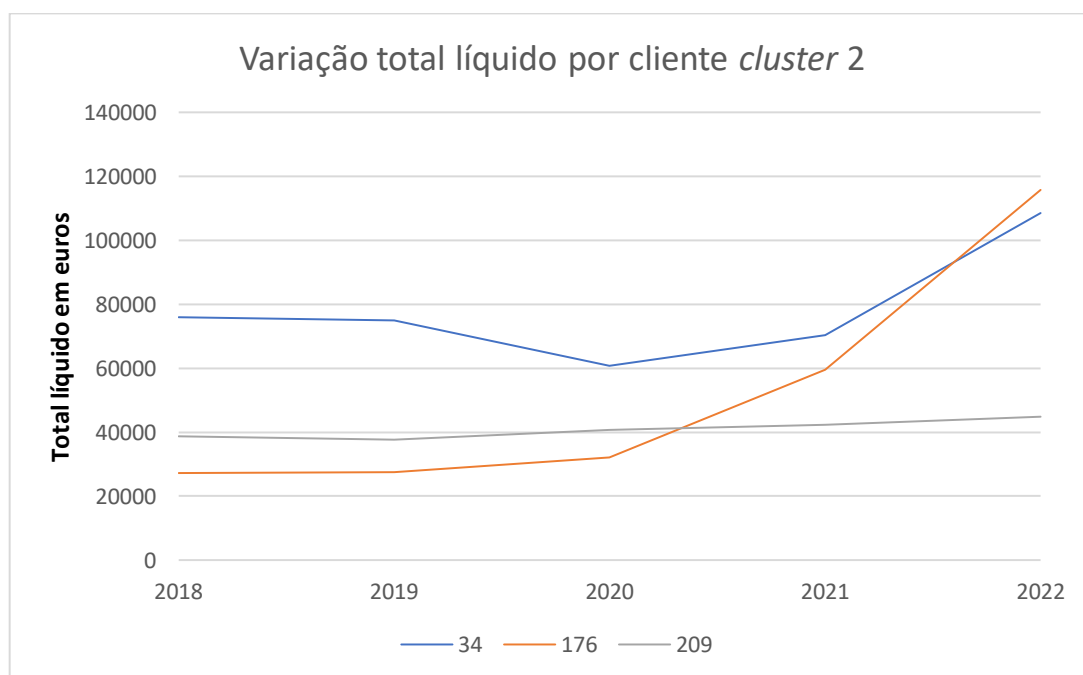


Figura 41 - Variação do total líquido por cliente do cluster 2

O Cliente 34 apresentou uma ligeira queda nas receitas entre 2018 e 2020, seguida de uma recuperação significativa em 2021 e 2022, atingindo o valor mais alto de 108.509 euros em 2022. O Cliente 176 teve um crescimento contínuo ao longo dos anos, com um aumento acentuado a partir de 2020, alcançando 115.746 euros em 2022, mostrando o crescimento mais expressivo. O Cliente 209 teve um crescimento lento com receitas de aproximadamente 44.832 euros em 2022, indicando uma tendência de crescimento estável e

menos volátil. Assim, o Cliente 176 destacou-se pelo crescimento mais consistente e significativo, enquanto o Cliente 34 mostrou uma recuperação robusta após um declínio inicial e o Cliente 209 manteve um crescimento estável ao longo dos anos.

A distribuição do tipo de peça mais faturada por cliente encontra-se representada na Figura 42. Observa-se que o cliente 34 apresenta maioritariamente peças do tipo POLO (60%) e T-SHIRT (33%), com pouco SWEAT (4%) e nenhum FELPA (0%). O cliente 176 caracteriza-se por uma maior percentagem de T-SHIRT (51%), com uma quantidade moderada de FELPA (15%) e pouco POLO (6%). O cliente 209 destaca-se pela predominância de T-SHIRT (69%), com quantidades menores de FELPA (4%) e POLO (11%) e um mínimo de SWEAT (1%). Estes dados indicam que as T-SHIRT são o tipo de peça mais comum nas encomendas, especialmente para os clientes 176 e 209, com POLO como segunda maior representatividade. Os restantes tipos de peças representam contributos menos significativos para o total líquido de cada um dos clientes do *cluster 2*.

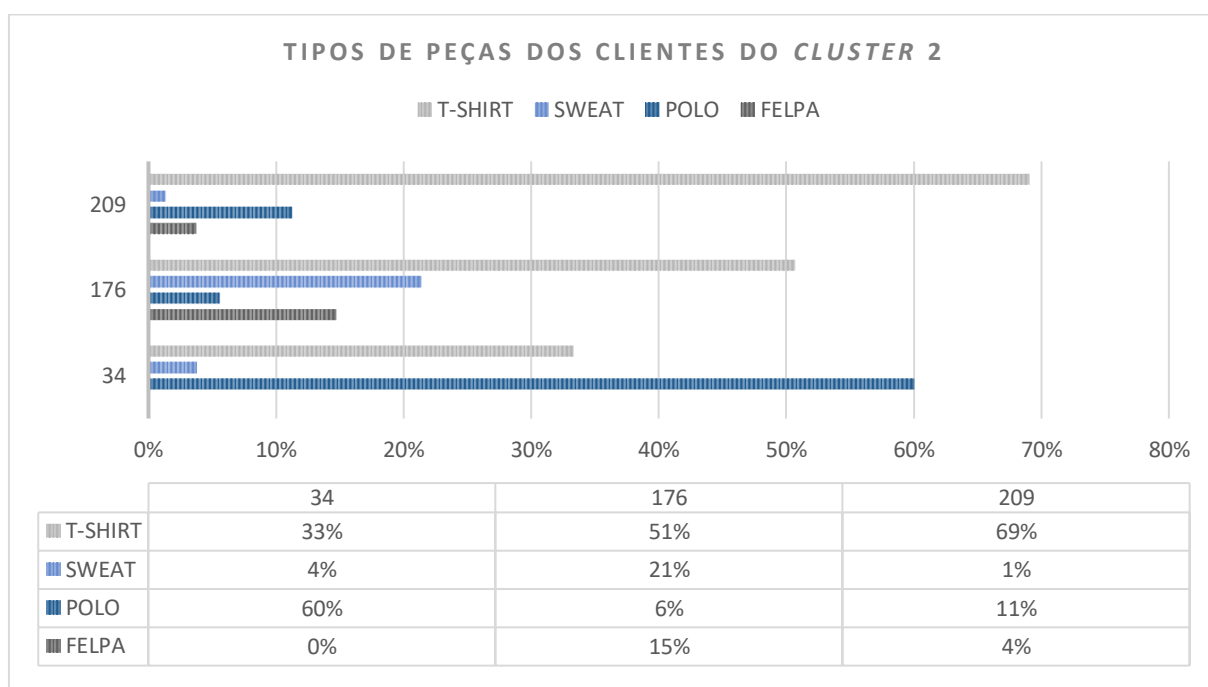


Figura 42 - Distribuição dos tipos de peça dos clientes do *cluster 2*

O tipo de lavagem preferencial de cada cliente pertencente ao *cluster 2*, encontra-se representado na Figura 43. O cliente 34 utiliza predominantemente a lavagem ESPECIAL 60° (80%), com uma menor frequência de lavagem NORMAL (15%) e quantidades muito baixas de lavagem C/DISPERSANTE (3%) e NORMAL C/SILICONE (1%). O cliente 176 opta principalmente pela lavagem NORMAL (74%), seguido pela lavagem NORMAL C/SILICONE (14%) enquanto as lavagens C/DISPERSANTE e ESPECIAL a 60° são menos utilizadas (5% e 1%, respetivamente). O cliente 209, por sua vez, mostra uma forte preferência pela lavagem

NORMAL C/SILICONE (63%), seguida pela lavagem C/DISPERSANTE (15%), enquanto as lavagens ESPECIAL 60° e NORMAL têm uma representação menor (2% e 11%, respetivamente). Estes dados indicam variações significativas nas preferências de lavagem entre os clientes, com destaque para o uso da lavagem ESPECIAL 60° pelo cliente 34, da lavagem NORMAL pelo cliente 176 e da lavagem NORMAL C/SILICONE pelo cliente 209.

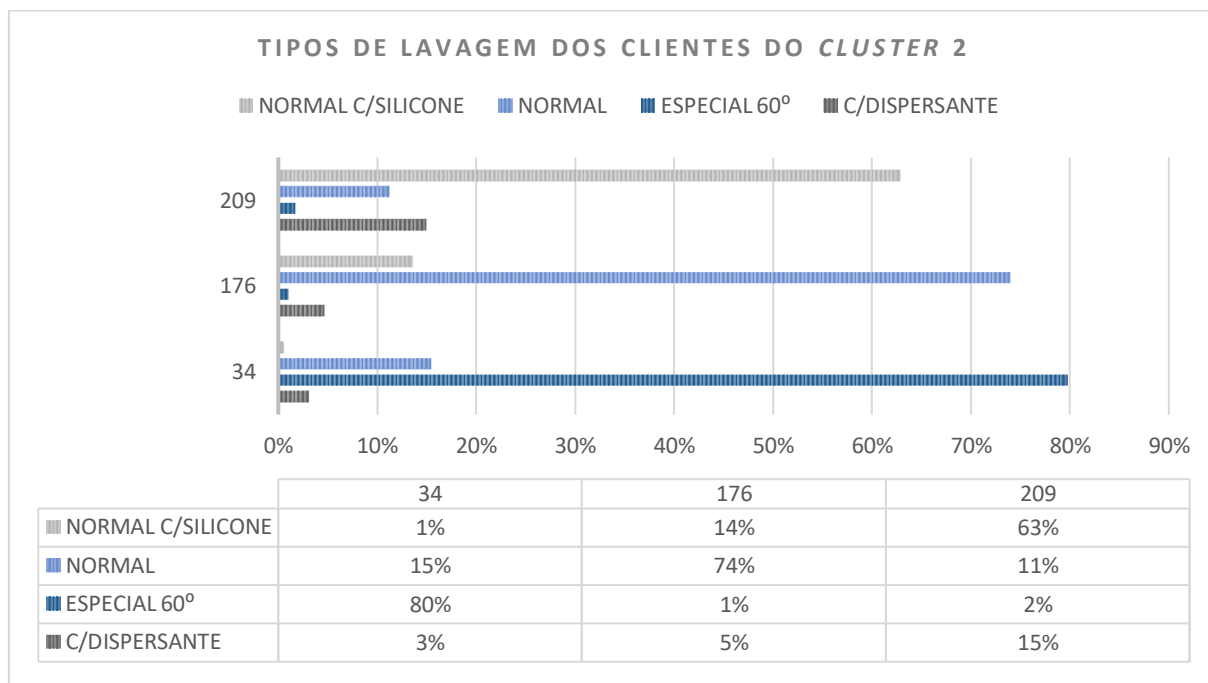


Figura 43 - Distribuição dos tipos de lavagem dos clientes do cluster 2.

Em resumo, a análise dos clientes do *cluster 2* revela que, apesar de representarem apenas 3,16% do total de clientes, eles contribuem com 47,564% do total de receitas entre 2018 e 2022. O Cliente 34 é o maior gerador de receita, seguido pelos Clientes 176 e 209. Cada cliente demonstra comportamentos distintos ao longo do período em análise, com o Cliente 176 a destacar-se pelo crescimento mais consistente e significativo, o Cliente 34 com uma recuperação forte após um declínio inicial e o Cliente 209 a manter crescimento estável.

Em termos de tipos de peças lavadas, o Cliente 34 coloca principalmente POLO e T-SHIRT, o Cliente 176 opta maioritariamente por T-SHIRT e FELPA e o Cliente 209 apresenta uma predominância de T-SHIRT seguido por POLO. Quanto aos tipos de lavagem, o Cliente 34 recorre preferencialmente a ESPECIAL 60°, o Cliente 176 prefere lavagem NORMAL e o Cliente 209 utiliza maioritariamente a lavagem NORMAL C/SILICONE.

6.2.3.2. Total líquido por tipo de lavagem

Relativamente ao total líquido por tipo de lavagem, conforme já referido anteriormente, o algoritmo que apresentou melhores resultados foi o k-médias, com distância

Euclidiana sem os dados estandardizados. Neste caso, o *clustering* obteve 8 *clusters*, sendo que o *cluster* 1 compreende 13 casos, representando 65 % dos 20 tipos de lavagem e todos os restantes sete *clusters* compreendem 1 caso cada, correspondendo a 5 % do total de tipos de lavagem.

A distribuição de cada um dos casos por *cluster* é a seguinte:

Tipo de Lavagem	Cluster	Tipo de Lavagem	Cluster
STONE-WASH	1	NORMAL	2
TUMBLER		ESPECIAL 60°	3
ENZIMÁTICA C/SILICONE		NORMAL C/SILICONE	4
ANTI-OLEO		SECO	5
ENZIMÁTICA		MICRONIZAR	6
60°		ANTI-PILLING	7
ANTI-PILLING C/SILICONE		C/DISPERSANTE	8
SAND WASH			
NORMAL AMOSTRAS			
PROGRAMA 40			
ANTI-ESTÁTICO			
ANTI-CORROSÃO			
FORTE			

Tabela 19 – Distribuição dos tipos de lavagem por cluster

Seguidamente é apresentado o comportamento da variação do total líquido por tipo de lavagem, ao longo do período de 5 anos em análise, ou seja, 60 meses, relativo ao *cluster* 1.

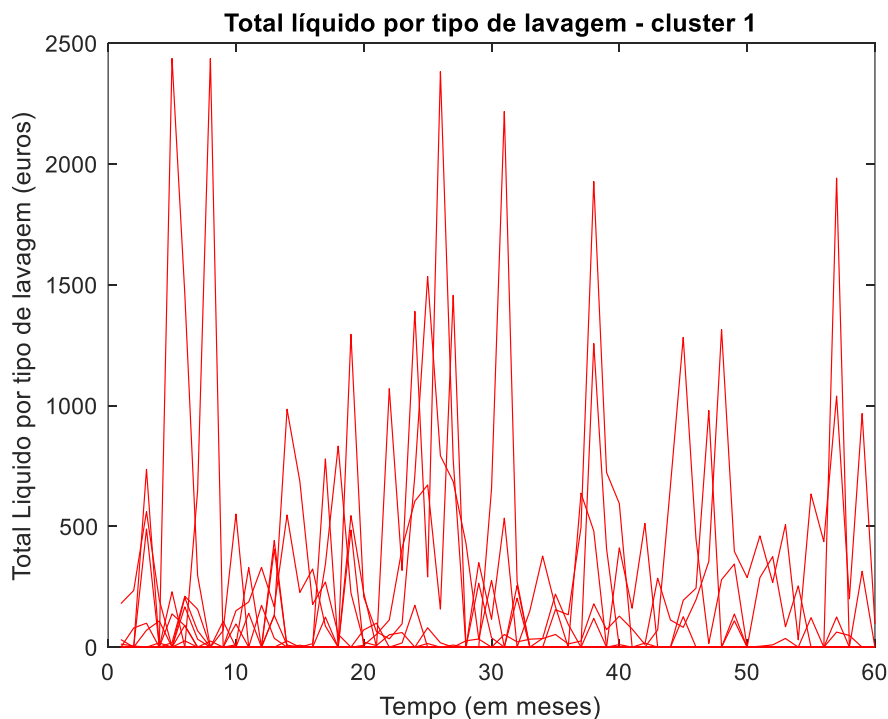


Figura 44 - Variação do total líquido por tipo de lavagem no cluster 1.

O total líquido acumulado destes 13 tipos de lavagem que constituem o *cluster* 1, correspondendo a 65% de total de tipos de lavagem, representa apenas 3,87 % do total líquido de todos os tipos de lavagem, ou seja, o *cluster* 1 engloba tipos de lavagem cujo contributo para o negócio da empresa é residual. Os diferentes tipos de lavagem no *cluster* 1 geram um total líquido baixo ao longo do tempo, embora, em alguns casos, com picos cujo valor máximo não ultrapassa os 2500 euros.

Os restantes sete *clusters* contêm cada um deles apenas um tipo de lavagem. Deste modo, estes sete tipos de lavagem são não só dissemelhantes entre si no que respeita ao total líquido, como também dissemelhantes em relação aos restantes tipos de lavagem no *cluster* 1.

Seguidamente é apresentado o comportamento da variação do total líquido por tipo de lavagem, ao longo do período em análise, relativo aos *clusters* 2 a 8.

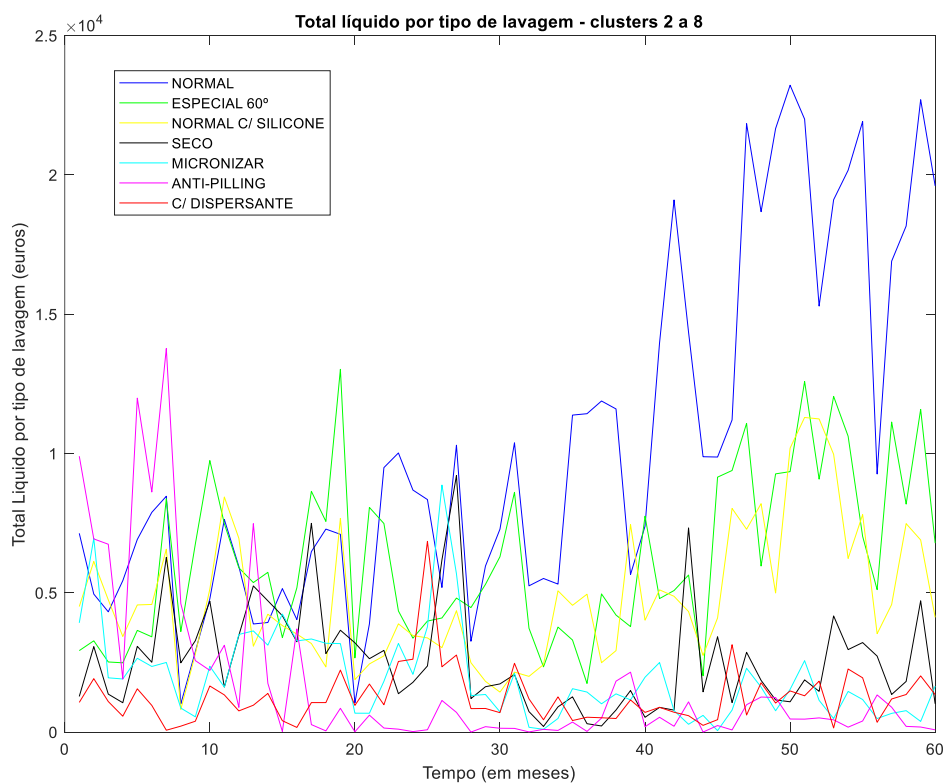


Figura 45 - Variação do total líquido por tipo de lavagem nos clusters 2 a 8.

O contributo para o negócio da empresa de cada um destes *clusters* varia entre o máximo de 34,05 %, correspondente ao tipo de lavagem NORMAL, até ao mínimo de 4,21 % para o tipo de lavagem C/DISPERSANTE. Desta forma, verifica-se que todos os tipos de lavagem que integram o *cluster* 1, contribuem em conjunto menos do que qualquer um dos tipos de lavagem que constituem os restantes oito *clusters*, individualmente. (Ver Figura 46)

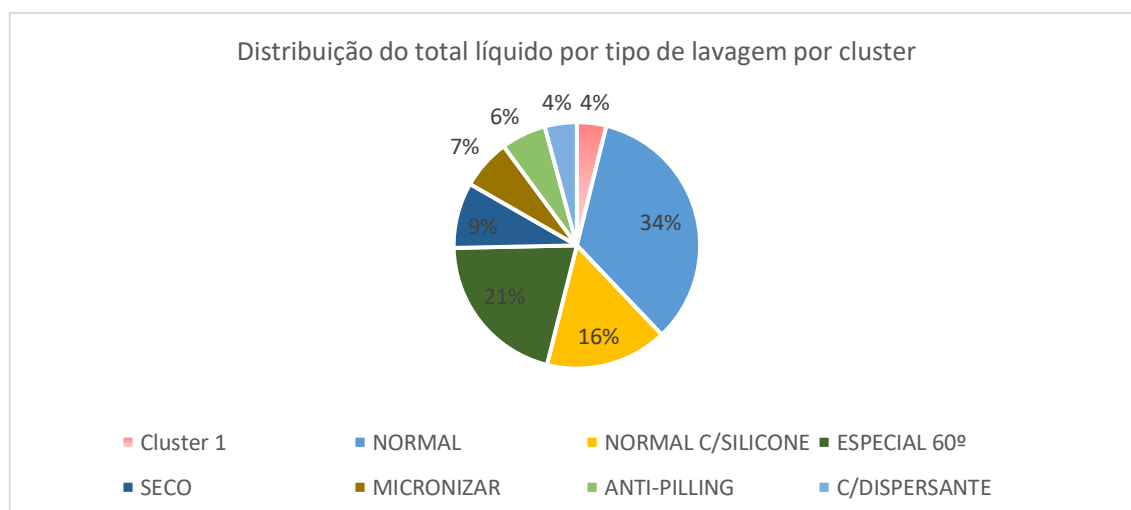


Figura 46 - Distribuição do total líquido por tipo de lavagem

A Figura 46 parece sugerir que, em geral, a dissemelhança entre os tipos de lavagem considerados se acentua na segunda metade do período em análise. É de realçar que, com o tempo, o total líquido das lavagens NORMAL, ESPECIAL 60° e NORMAL C/ SILICONE tendeu a ser mais elevado e a crescer de forma significativa no caso da lavagem NORMAL. Em contraponto, o ANTI-PILLING tendeu a perder importância em termos de volume de negócio. Com o tempo, o total líquido desta lavagem e das lavagens SECO, do tipo MICRONIZAR e C/ DISPERSANTE tendeu a manter-se baixo e mais estável.

Na figura seguinte está representado o total líquido por tipo de lavagem, dos 8 *clusters*, sendo que no caso do *cluster 1* a representação é a da média dos diferentes tipos de lavagem que constituem o *cluster*.

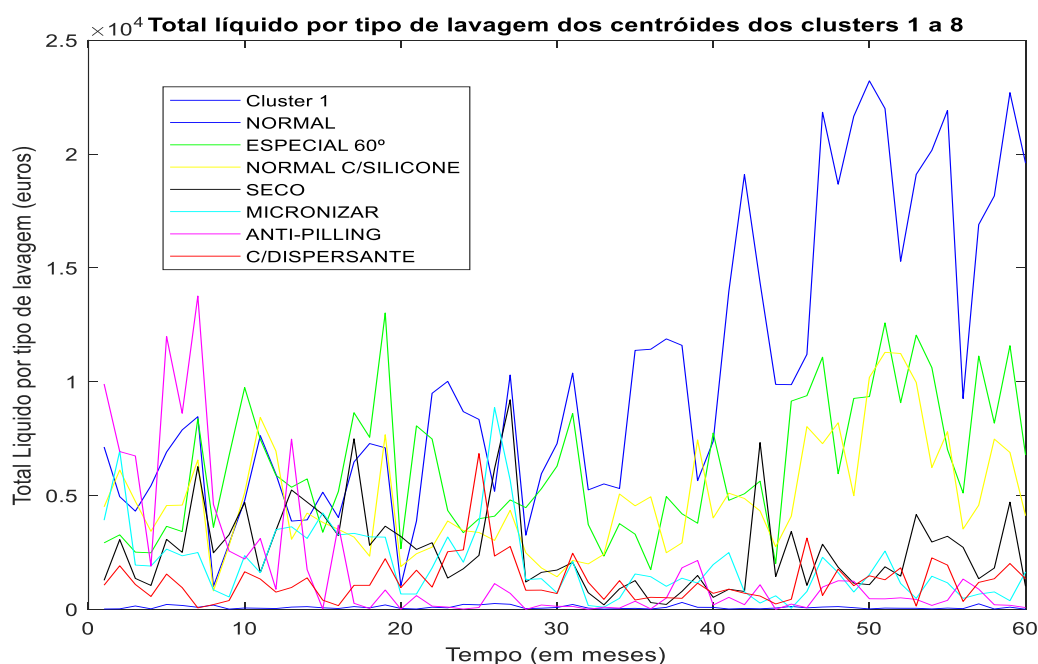


Figura 47 - Variação do total líquido por tipo de lavagem dos clusters 1 a 8.

Da análise da Figura 47 é possível verificar que a média do *cluster* 1 (representada a azul escuro, em baixo) é inferior ao que se regista nos outros *clusters*.

De seguida é apresentada a distribuição do tipo de lavagens no *cluster* 1.

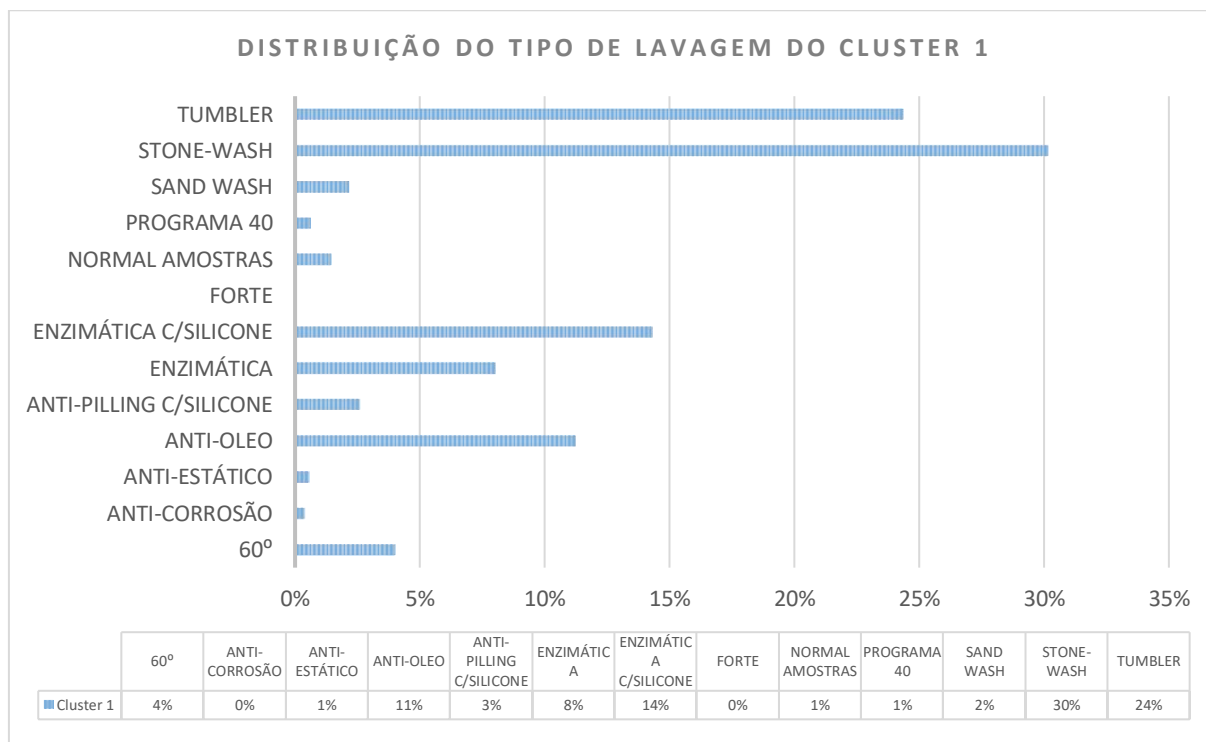


Figura 48 - Distribuição do tipo de lavagens integrantes do *cluster* 1.

A análise do gráfico anterior, revela que a lavagem mais comum é a STONE-WASH, representando 30% do total, seguida pela TUMBLER com 24% e pela ENZIMÁTICA C/SILICONE com 14%. A lavagem ANTI-OLEO aparece com 11%, enquanto a ENZIMÁTICA corresponde a 8% das lavagens. Lavagens menos frequentes incluem a 60° (4%), ANTI-PILLING C/SILICONE (3%) e SAND WASH (2%). Vários tipos de lavagem, como ANTI-CORROSÃO e FORTE praticamente não são utilizadas

Seguidamente é apresentado um gráfico que indica a distribuição do tipo de peça por tipo de lavagem que constituem o *cluster* 1. Para possibilitar uma melhor perceção dos dados, apenas foram considerados tipos de lavagem e tipos de peça que representem mais de 3% do total. Note-se que a análise completa envolveria considerar os 8 tipos de lavagem do *cluster* 1 e os 43 tipos de peça existentes, mas o contributo de muitas das combinações possíveis é marginal ou nulo para o negócio da empresa.

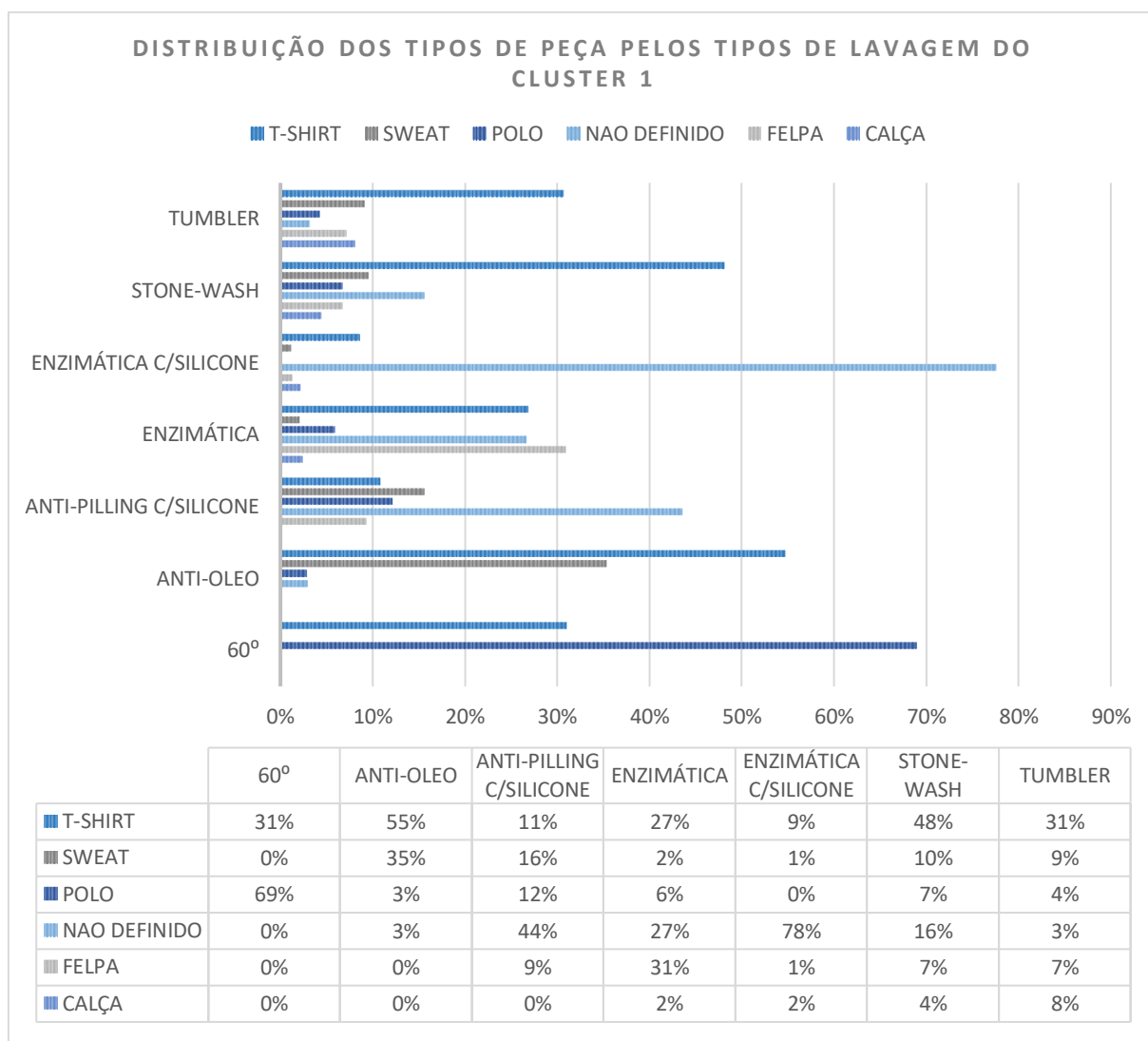


Figura 49 - Distribuição dos tipos de peça por tipos de lavagens integrantes do cluster 1.

Da análise da Figura 49 verifica-se que para o tipo de peça T-SHIRT, os tipos de lavagem integrantes do *cluster 1* mais utilizados são o ANTI-OLEO (55%), STONE-WASH (48%), TUMBLER e 60° (ambos com 31%). No tipo POLO, a lavagem deste *cluster* mais aplicada é a 60° com 69%, enquanto outros tipos, como ANTI-PILLING C/SILICONE e STONE-WASH, têm valores menores (12% e 7%, respetivamente). No tipo de peça de FELPA destaca-se a lavagem ENZIMÁTICA (31%), seguida de ANTI-PILLING C/SILICONE (9%), TUMBLER e STONE-WASH (com 7%). O tipo SWEAT têm uma alta percentagem na lavagem ANTI-OLEO (35%), um de valor considerável no ANTI-PILLING C/SILICONE (16%) e com menor expressão o STONE-WASH (10%). Para o tipo de peça NÃO DEFINIDO as lavagens ENZIMÁTICA C/SILICONE e ANTI-PILLING C/SILICONE têm os valores mais altos, com 78% e 44%, respetivamente, destacando-se ainda o tipo ENZIMÁTICA com 27%. Finalmente, no tipo CALÇA as lavagens mais frequentes são TUMBLER (8%) e STONE-WASH (4%), enquanto os restantes tratamentos registam percentagens muito baixas ou nulas.

O próximo gráfico indica a distribuição dos tipos de peça pelos *clusters* 2 a 8 (C2 a C8, correspondendo cada um a um tipo de lavagem). Tal como no gráfico anterior, para possibilitar uma melhor percepção dos dados, apenas foram considerados tipos de lavagem e tipos de peça que representem mais de 3% do total.

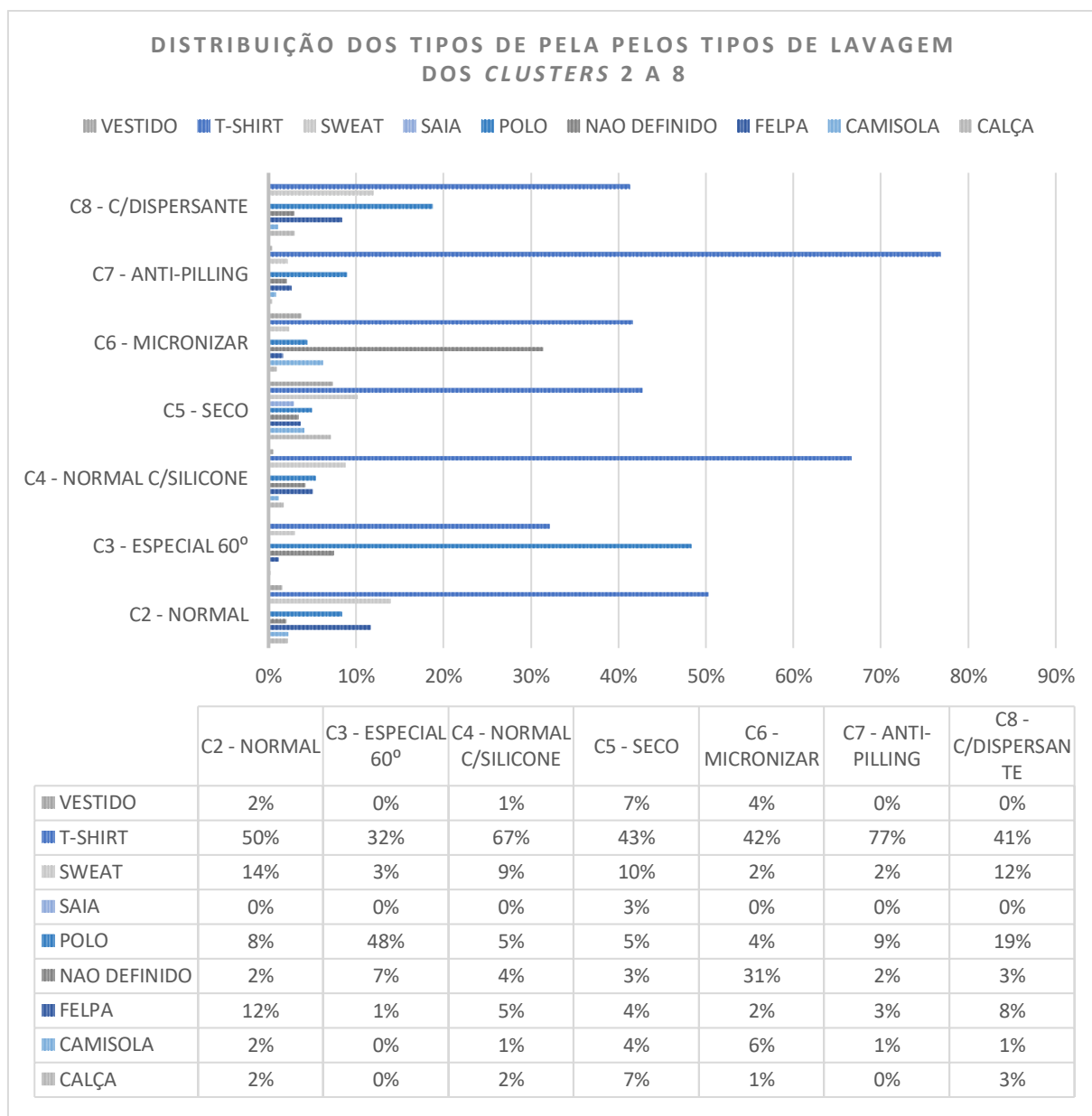


Figura 50 - Distribuição dos tipos de peça por tipos de lavagens integrantes dos *clusters* 2 a 8.

Da análise da Figura 50 pode-se verificar que a lavagem **NORMAL (C2)** é mais aplicada em T-SHIRT (50%) e SWEAT (14%), com percentagens menores em FELPA (12%) e POLO (8%). As restantes categorias, como CALÇAS e CAMISOLAS, têm valores baixos (2% cada). A lavagem **ESPECIAL 60° (C3)** é predominantemente utilizada em POLO (48%) e, em menor medida, em T-SHIRT (32%). Outras peças, como FELPA (1%) e SWEAT (3%), quase não utilizam este tipo de lavagem. A lavagem **NORMAL C/SILICONE (C4)** apresenta

uma alta percentagem no tipo T-SHIRT (67%), seguida por SWEAT (9%), POLO e FELPA (ambos com 6%). Os restantes tipos de peça registam valores mais reduzidos. A lavagem a **SECO (C5)** é aplicada principalmente em T-SHIRT (43%), seguida por SWEAT (10%) e VESTIDOS e CALÇA (ambos com 7%). Os restantes tipos têm percentagens mais baixas. A lavagem **MICRONIZAR (C6)** tem um uso significativo em T-SHIRT (42%) e no tipo de peça NÃO DEFINIDO (31%). Os tipos CAMISOLA e POLO representam valores pouco significativos (6% e 4%, respetivamente), com os restantes tipos de peça a terem baixas percentagens. A lavagem **ANTI-PILLING (C7)** é mais frequente em T-SHIRT (77%) e POLO (9%) e praticamente ausente nas outras categorias, exceto com pequenas percentagens em SWEAT (2%), NÃO DEFINIDO (2%) e FELPA (3%). Por fim, o tipo de lavagem **C/DISPERSANTE (C8)** é mais comum em T-SHIRT (41%) e POLO (19%), seguido por SWEAT (12%) e FELPA (8%). Os restantes tipos de peça, apresentam valores muito reduzidos.

Em resumo, a análise dos tipos de lavagem com o algoritmo k-médias revelou que o *cluster* 1, composto por 13 tipos de lavagem, representa 65% dos tipos existentes, mas contribui apenas com 3,87% do total líquido, indicando baixa relevância para o mesmo. Por contraste, os *clusters* 2 a 8, com apenas um tipo de lavagem cada, contribuem de forma significativa para o negócio, especialmente a lavagem NORMAL com 34,05% do total líquido. As lavagens NORMAL, ESPECIAL 60° e NORMAL C/SILICONE revelaram um crescimento ao longo do tempo, enquanto outros tipos de lavagens como ANTI-PILLING, perderam volume de negócio.

6.2.3.3. Total líquido por tipo de peça

Relativamente ao total líquido por tipo de peça, conforme já foi referido antes, o algoritmo que apresentou melhores resultados foi o k-médias, com distância Euclidiana sem os dados estandardizados. Neste caso, o *clustering* obteve 2 *clusters*, em que o *cluster* 1 compreende 42 casos, representando 97.67 % do total de peças, sendo que o *cluster* 2 compreende 1 caso, o tipo de peça T-SHIRT, correspondendo a 2,33 % do total de peças.

Tipo de Peça			
AMOSTRAS	CORDÃO	MALHAS	SAIA
BLAZER	CORTINAS	MÁSCARAS	SINGLET
BLUSA	CREPOM	MEIAS	SUJOS
BODY	CUECA	NAO DEFINIDO	SWEAT
CALÇA	FELPA	PAINEIS	TIRAS
CALÇÃO	FITAS	PANOS	TOP
CAMISA	GOLAS	PIJAMAS	T-SHIRT
CAMISOLA	HOODIES	POLO	TUBO
CAPA	LEGGINS	PUFS	TUNICA
CASACO	LENÇOS	RENDAS	VESTIDO
COLETES	MACACÃO	ROBE	

Tabela 20 - Distribuição dos tipos de peça por cluster. A vermelho o cluster 1 e a azul o cluster 2.

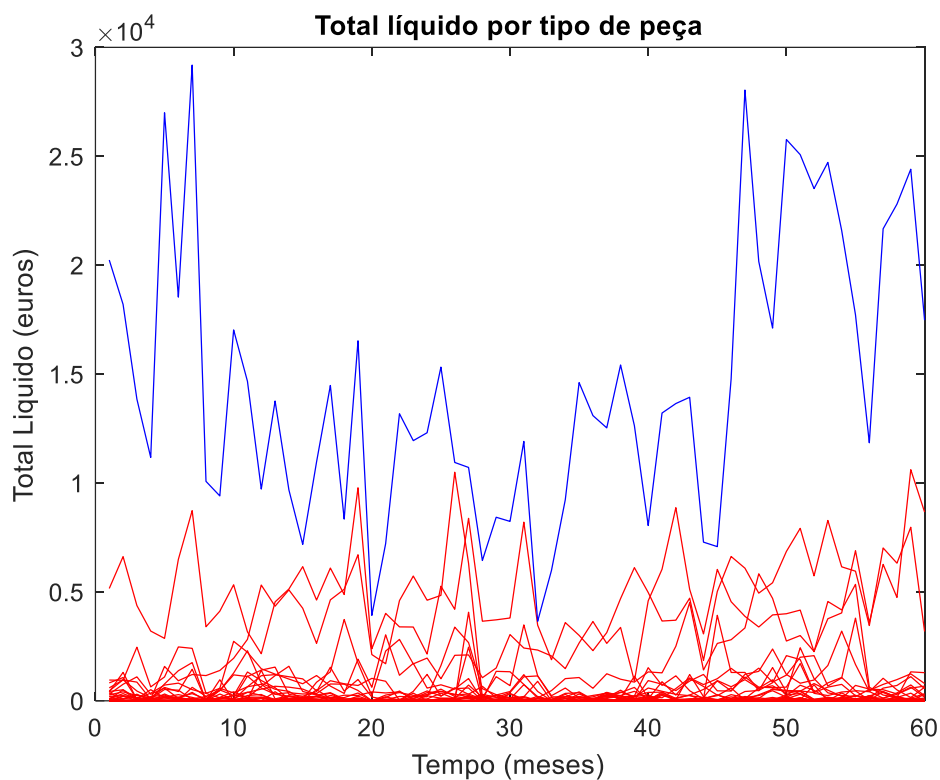


Figura 51 - Variação do total líquido por tipo de peça (cluster 1 a vermelho, cluster 2 a azul).

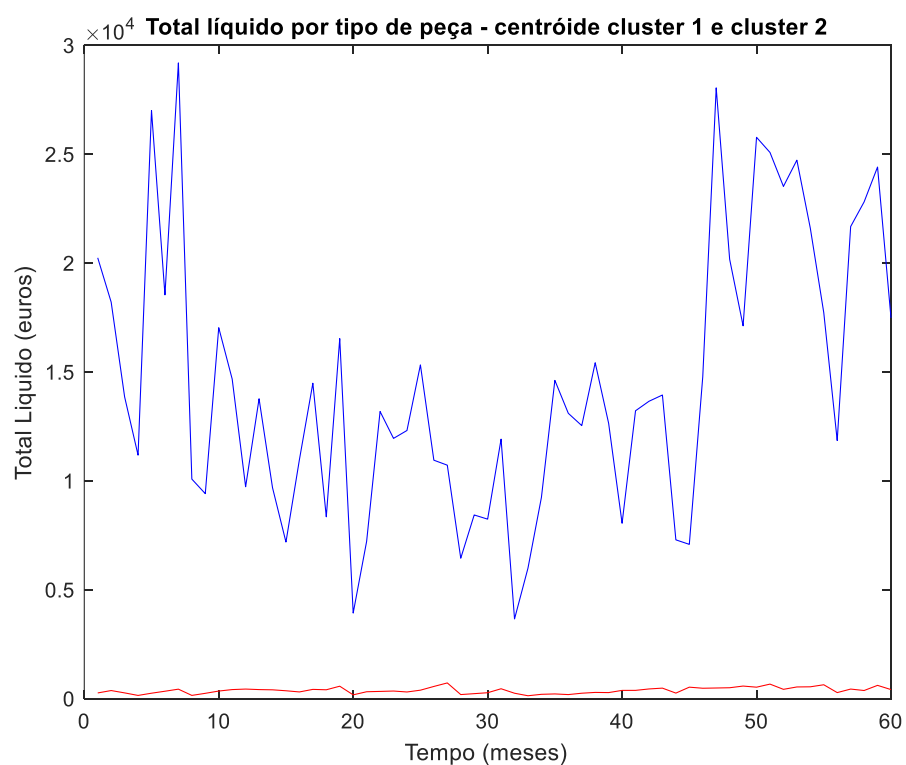


Figura 52 - Variação do total líquido por tipo de peça (centroide do cluster 1 a vermelho, cluster 2 a azul).

A análise dos gráficos anteriores permite verificar que o total líquido gerado pela peça T-SHIRT, a única no *cluster 2*, foi sempre superior ao total líquido gerado pelas restantes

peças, todas elas no *cluster* 1. Isto sucedeu mesmo quando o volume de negócio com a T-SHIRT diminuiu. Conforme é visível, o total líquido desta peça não foi estável ao longo do tempo. De facto, tendeu a diminuir e depois a aumentar. Em oposição, o total líquido das peças no *cluster* 1 tendeu a ser estável.

No período em análise, o total líquido respeitante à T-SHIRT é 48,17 % do total líquido global. Outros 37,4 % do volume de negócio estão concentrados em quatro tipos de peça no *cluster* 1, nomeadamente POLO, SWEAT, FELPA e NÃO DEFINIDO. Assim, as restantes trinta e oito peças, isto é, 88,37 % das 43 existentes, explicam apenas 14,46 % da faturação da empresa.

O gráfico seguinte representa o total líquido anual por tipo de peça relativamente aos quatro tipos de peça do *cluster* 1 referidos no ponto anterior. Assim, podemos verificar que o tipo de peça POLO é o que tem um comportamento mais uniforme, com tendência de subida mais acentuada a partir de 2021. Relativamente a FELPA, verifica-se uma subida significativa até 2021, passando depois para um período de descida também acentuada. Quanto à SWEAT, a partir de 2020 regista uma subida muito relevante, terminando o período em análise com valores de total líquido próximos ao POLO. O tipo de peça NÃO DEFINIDO apresenta uma variação significativa, sobre a qual não é possível elaborar muito mais dado o desconhecimento do tipo de peça.

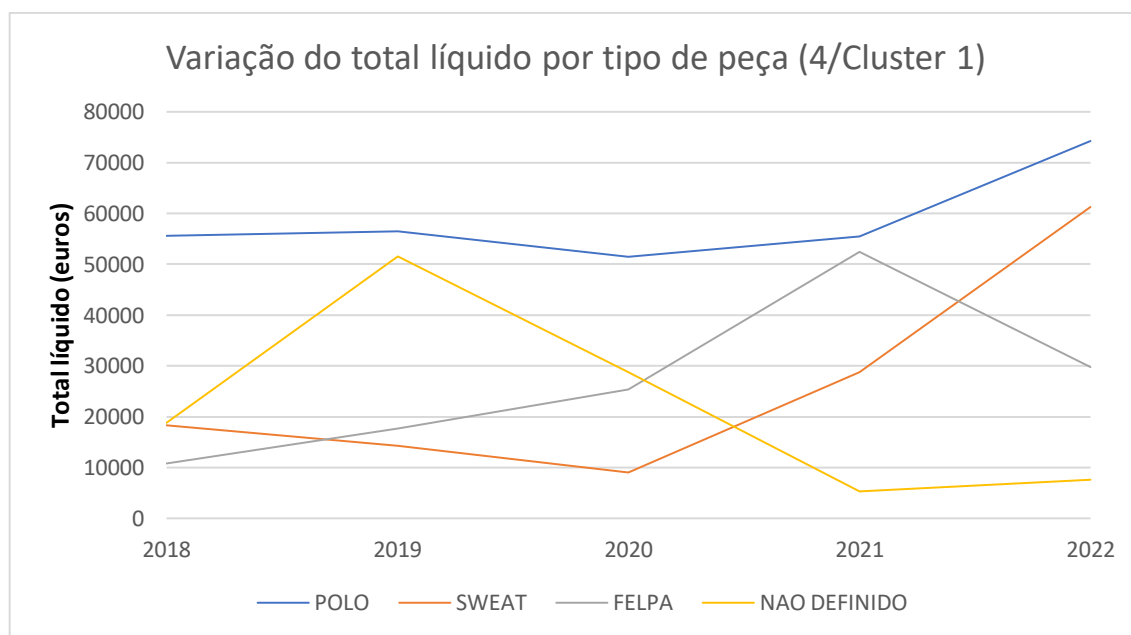


Figura 53 - Variação do total líquido por tipo de peça.

O gráfico seguinte indica a distribuição dos tipos de lavagem por tipo de peça integrantes do *cluster*. De forma a permitir uma melhor perceção dos dados apenas são considerados os tipos de peças e de lavagem que correspondem a mais de 3% do total.

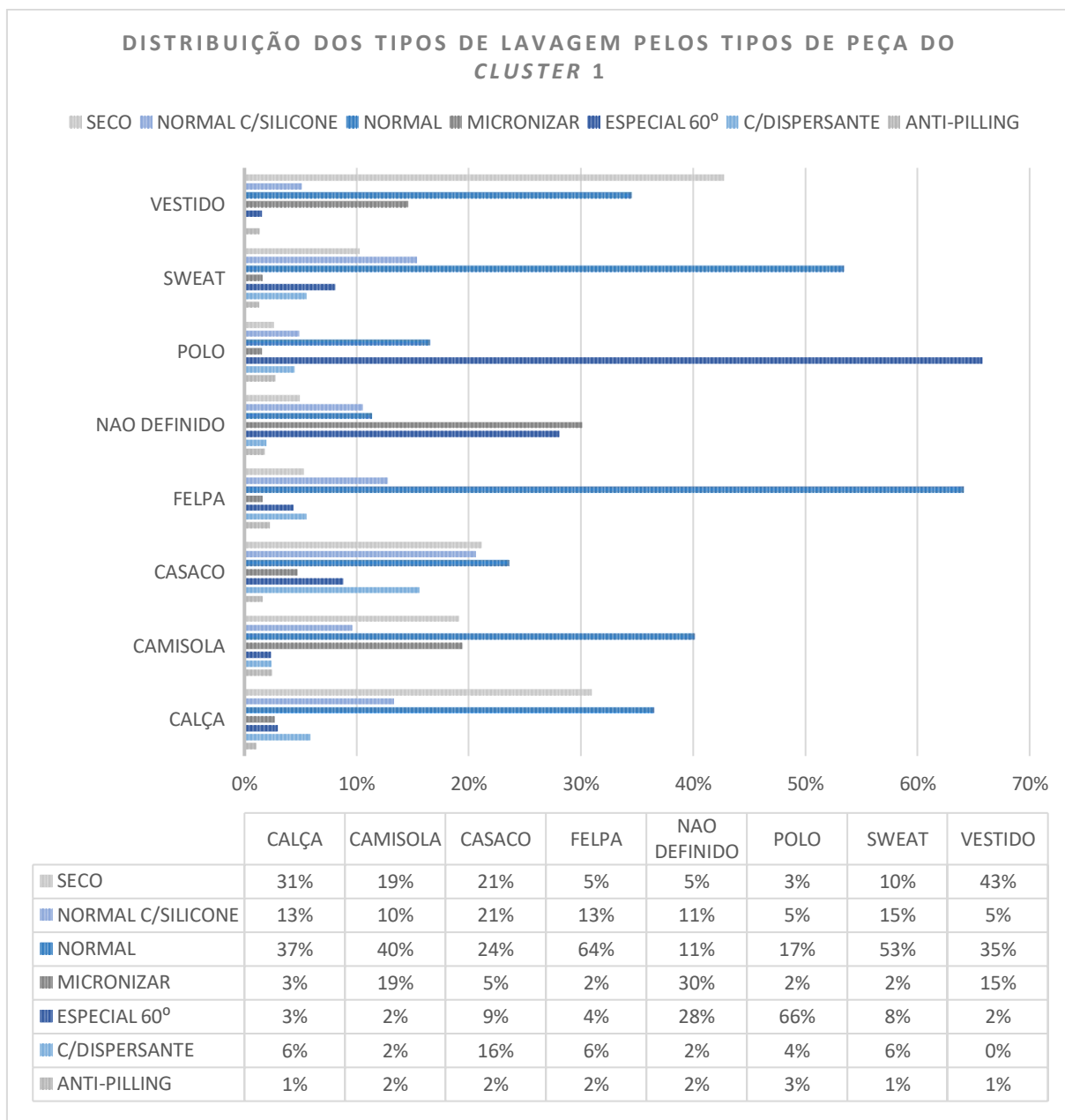


Figura 54 - Distribuição dos tipos de lavagem pelos tipos de peça do cluster 1.

O tipo de peça **CALÇA** é predominantemente lavadas com o tipo de lavagem NORMAL (37%) e SECO (31%). A lavagem NORMAL C/SILICONE também é relativamente comum (13%), enquanto outras lavagens, como MICRONIZAR (3%) e ESPECIAL 60º (3%) têm baixa expressão. Quanto ao tipo **CAMISOLA** a lavagem NORMAL é a mais utilizada (40%), seguida por SECO (19%) e MICRONIZAR (19%). O uso de NORMAL C/SILICONE (10%) é mais reduzido, com os restantes tipos de lavagem com baixa expressão. O tipo de peça **CASACO** tem maior expressão no tipo de lavagem NORMAL (24%), com percentagens significativas para SECO (21%) e NORMAL C/SILICONE (21%). A lavagem C/DISPERSANTE é também significativa (16%), bem como a ESPECIAL 60º. O tipo **FELPA** tem uma grande utilização do tipo de lavagem NORMAL (64%), seguido por NORMAL C/SILICONE (13%) e

C/DISPERSANTE (6%). Os restantes tipos de lavagem, apesar de serem utilizados, têm uma baixa expressão. O tipo de peça **NÃO DEFINIDO** utiliza maioritariamente os tipos de lavagem ESPECIAL 60° (28%) e MICRONIZAR (30%), seguido pelos tipos NORMAL e NORMA C/SILICONE (ambos com 11%), tendo os restantes tipos de lavagem baixa relevância. Quanto ao tipo de peça **POLO** a lavagem ESPECIAL 60° domina (66%), seguido pelo tipo NORMAL (17%), sendo que os restantes tipos têm pouca expressão. No que concerne ao tipo de peça **SWEAT**, os tipos de lavagem mais utilizados são NORMAL (53%), NORMAL C/SILICONE (15%) e SECO (10%), com os restantes tipos a terem valores menores e menos relevantes. Por fim, o tipo de peça **VESTIDO** tem como lavagem mais comum o SECO (43%), seguido dos tipos NORMAL (35%) e MICRONIZAR (15%). Todos os outros tipos de lavagem têm uma utilização menor neste tipo de peça.

Relativamente ao *cluster 2*, constituído apenas pelo tipo de peça T-SHIRT, é seguidamente apresentada a distribuição dos tipos de lavagem dentro deste *cluster*.

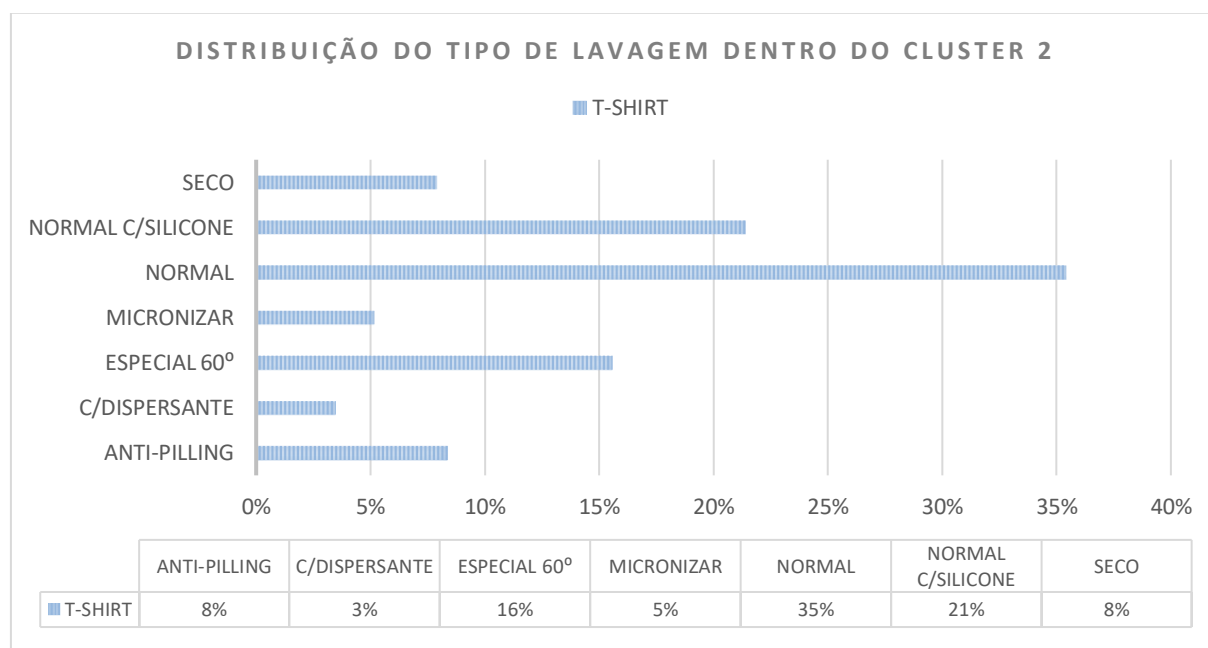


Figura 55 - Distribuição dos tipos de lavagem dentro do tipo de peça do *cluster 2*.

O tipo de peça T-SHIRT é predominantemente lavado com a lavagem NORMAL (35%) e NORMAL C/SILICONE (21%). Outras lavagens com significância são a ESPECIAL 60° (16%), SECO e ANTI-PILLING (ambos com 8%). Os tipos de lavagem MICRONIZAR (5%) e C/DISPERSANTE (3%) são menos utilizadas neste tipo de peça.

Os clientes mais representativos ao nível do tipo de peça T-SHIRT, correspondente ao *cluster 2*, são apresentados na seguinte figura:

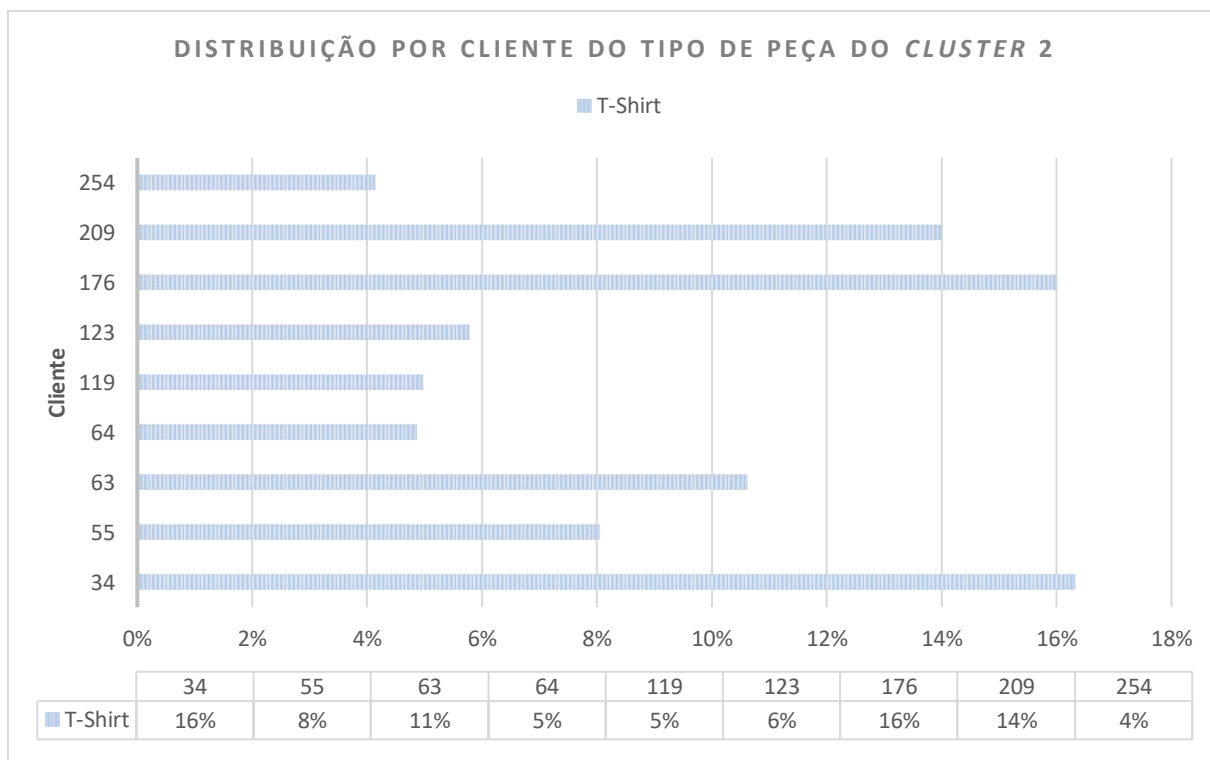


Figura 56 - Distribuição por clientes do tipo de peça do cluster 2 (T-SHIRT).

A análise da distribuição percentual do tipo de peça T-SHIRT revela que os clientes 34 (16%), 176 (16%) e 209 (14%) têm a maior proporção. De seguida vem o cliente 63 com uma percentagem de 11%. Os restantes clientes têm uma menor expressão, variando entre os 4% e os 8%.

Esta análise permite verificar que os clientes que constituem o *cluster* 2, relativamente aos *clusters* do total líquido por cliente (34, 176 e 209), são também os mais representativos ao nível do tipo de peça que constitui o *cluster* 2 quanto ao total líquido por tipo de peça.

Em resumo, a análise identificou 2 *clusters* distintos no total líquido por tipo de lavagem. O *cluster* 1 inclui 42 tipos de peças, representando 97,67% do total de peças, enquanto o *cluster* 2 é composto exclusivamente pela peça T-SHIRT, correspondendo a 2,33% do total de peças.

O tipo de peça T-SHIRT, apesar de ser menos estável ao longo do tempo, em termos de faturação, gerou 48,17% do total líquido global, destacando-se sobre os outros tipos de peças do *cluster* 1, que incluem POLO, SWEAT, FELPA e peças NÃO DEFINIDAS, responsáveis por 37,4% do total líquido. Os tipos de lavagem mais frequentes para T-SHIRT são NORMAL e NORMAL C/SILICONE.

Para os tipos de peça do *cluster* 1, as lavagens variam conforme o tipo de peça, nos tipos CALÇA e CAMISOLA predominam nas lavagens NORMAL e SECO, nos tipos CASACO e SWEAT destacam-se com NORMAL, NORMAL C/SILICONE, e SECO, quanto ao tipo de

peça FELPA, este tem como tipo de lavagem mais significativo o NORMAL. O tipo NÃO DEFINIDO tem como mais significativos os tipos de lavagem ESPECIAL 60° e MICRONIZAR. O tipo POLO utiliza maioritariamente o tipo de lavagem ESPECIAL 60° e o tipo VESTIDO privilegia SECO e NORMAL

A análise também revelou que os clientes 176, 209, e 34 são os principais contribuidores para o total líquido gerado pela peça T-SHIRT.

6.3. Previsão de séries temporais

Conforme já foi referido, os dados de faturação recolhidos no âmbito deste estudo podem ser organizados em séries temporais, tendo em conta que a cada registo corresponde determinada data.

Com o objetivo de contribuir para a orientação da empresa para uma gestão baseada em dados, torna-se relevante fazer previsões a partir da informação disponível, no sentido de estimar dados futuros sobre o desenvolvimento da atividade da lavandaria em consideração.

Assim, depois de apresentados os modelos de previsão de séries temporais, estes são aplicados ao presente estudo. No entanto, tendo em conta a limitação em termos de extensão temporal deste trabalho, a análise preditiva é restringida à evolução dos valores mensais do total líquido.

Os dados recolhidos no período de 2018 a 2022, apresentam relativamente ao total líquido a seguinte distribuição gráfica:

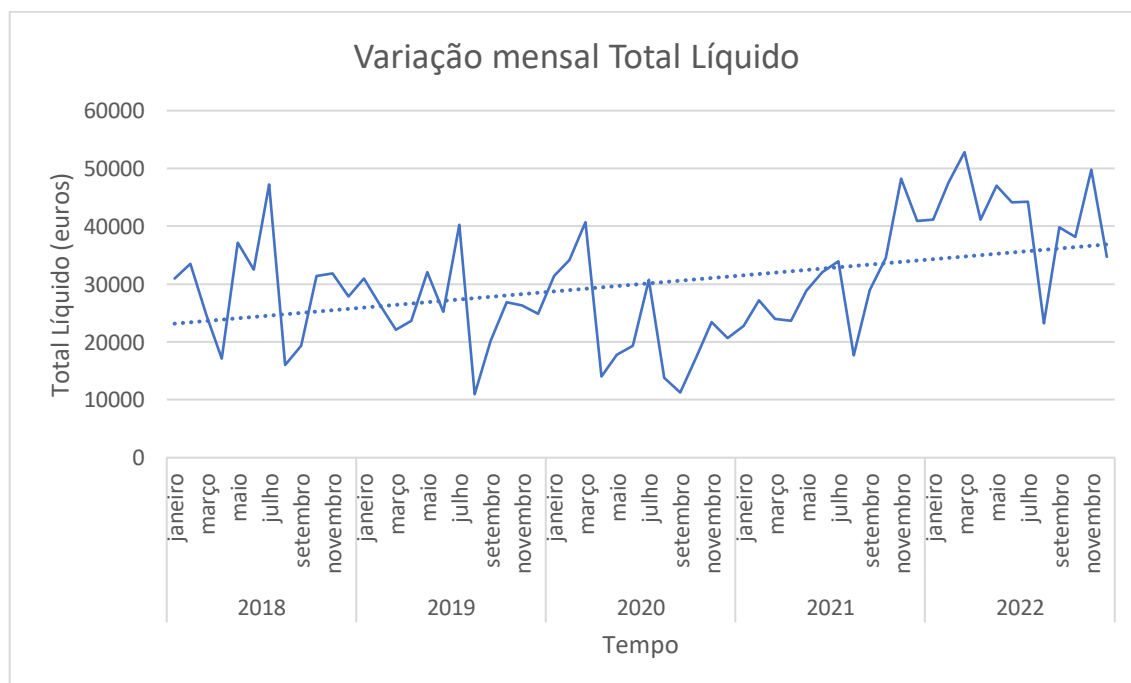


Figura 57 - Representação gráfica da variação do total líquido no período em análise

Como se pode verificar, existem oscilações mensais significativas, com picos e quedas acentuadas. Apesar das flutuações mensais, é evidenciado um crescimento gradual ao longo do tempo. Os dados permitem ainda observar um possível padrão sazonal, com variações recorrentes que podem estar relacionadas com ciclos anuais ou trimestrais. Este comportamento sugere a presença tanto de uma tendência de longo prazo como de elementos sazonais que influenciam os dados.

Em termos de tendência apresenta um crescimento gradual ao longo do tempo, verificando-se um padrão sazonal claro, com oscilações mensais que se repetem todos os anos e relativamente a resíduos apresenta flutuações relativamente controladas, com alguns desvios pontuais.

6.3.1. Total líquido global

Aos dados globais foram aplicados os oito modelos de previsão diferentes, ajustados de 2018 a 2022 e usados para previsão de 2023, tendo sido obtidos os seguintes resultados:

6.3.1.1. Método Mean

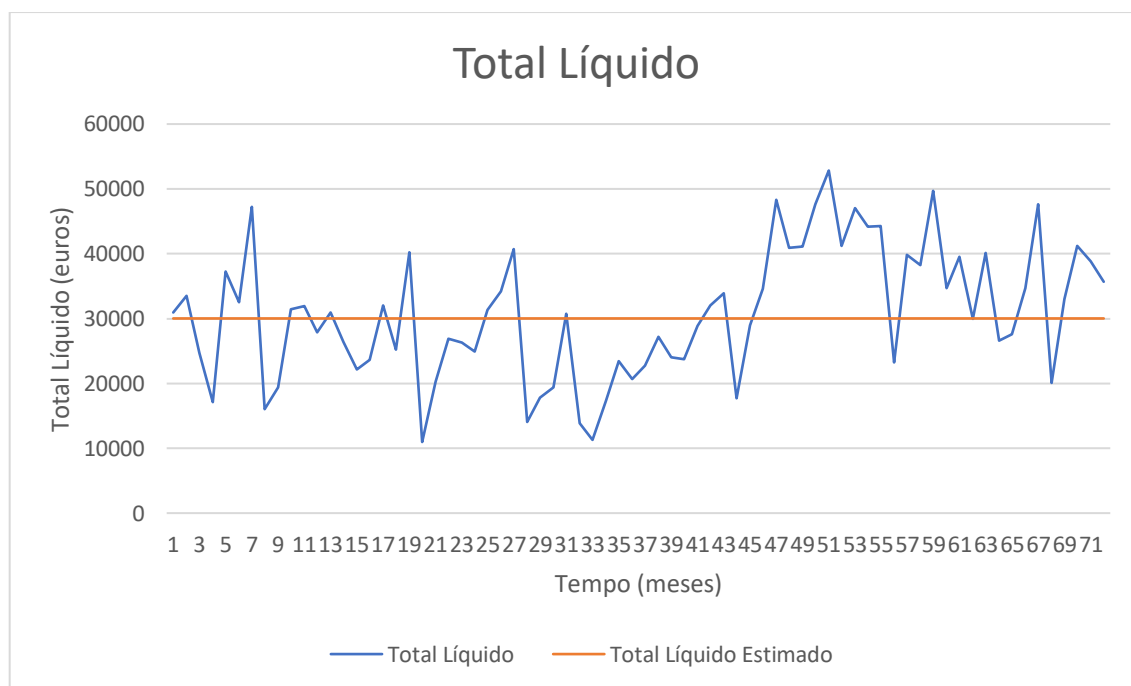


Figura 58 - Representação gráfica da aplicação do método média à estimativa do total líquido

6.3.1.2. Método Naive

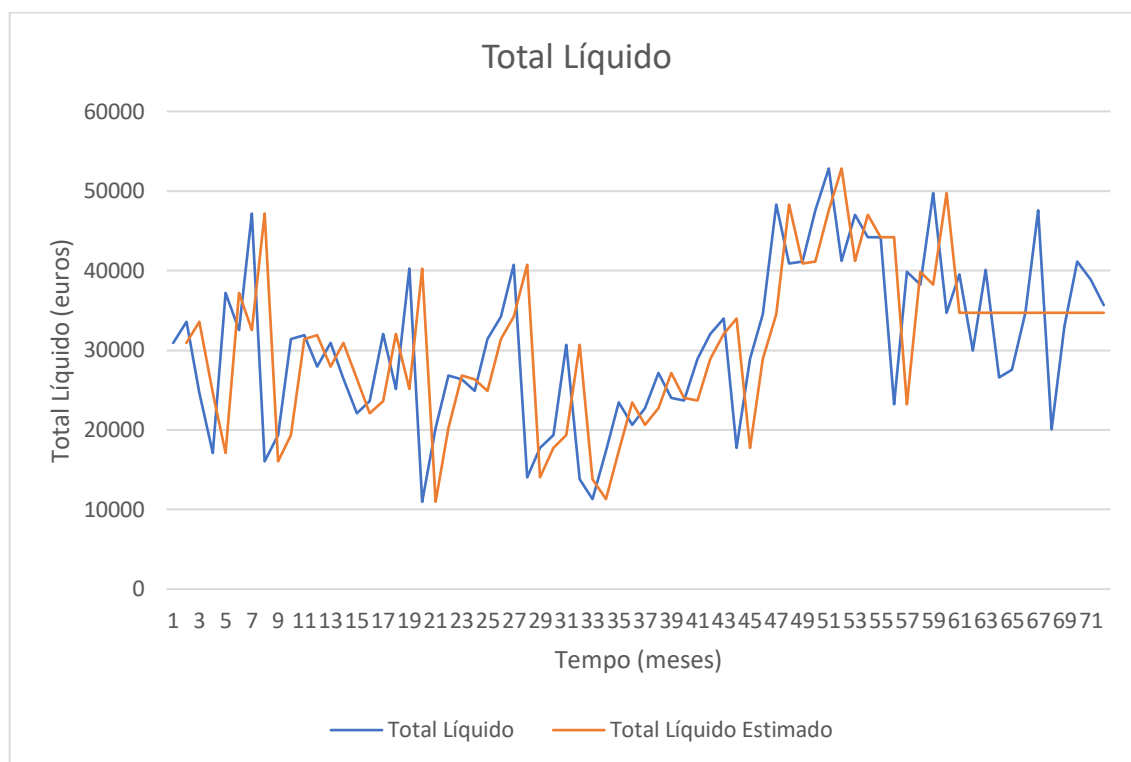


Figura 59 -Representação gráfica da aplicação do método ingénuo à estimativa do total líquido

6.3.1.3. Método SNaive

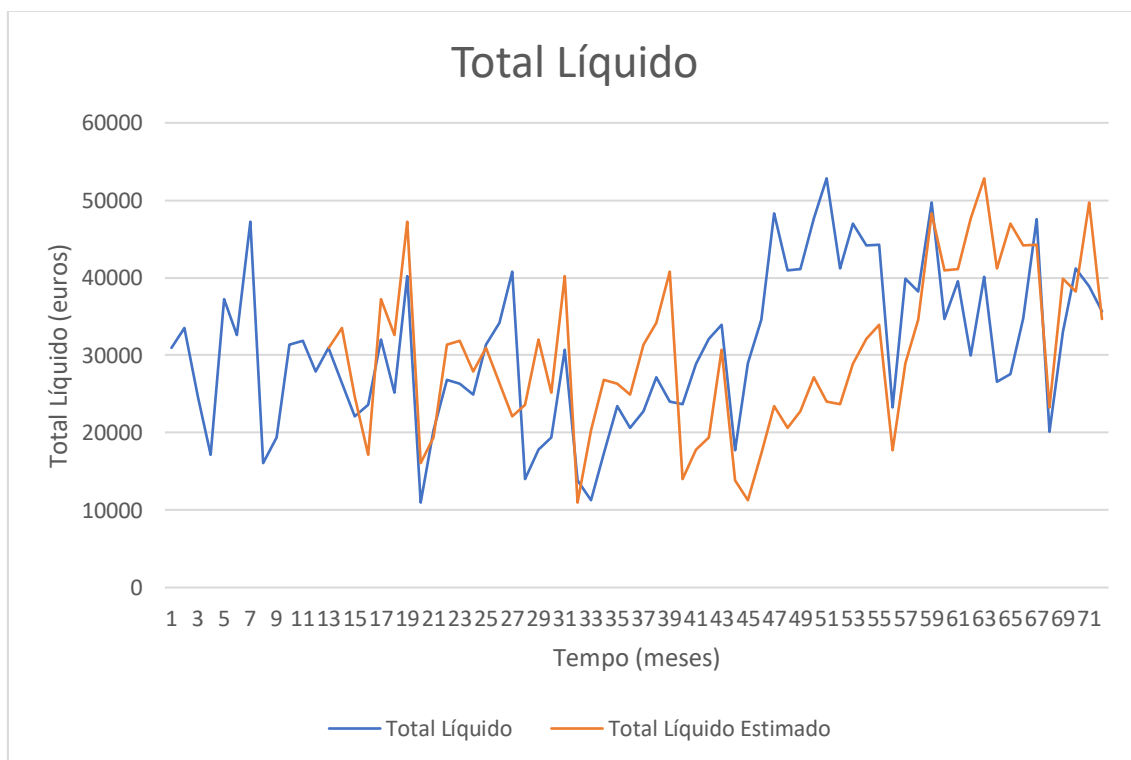


Figura 60 - Representação gráfica da aplicação do método ingénuo sazonal à estimativa do total líquido

6.3.1.4. Método Drift

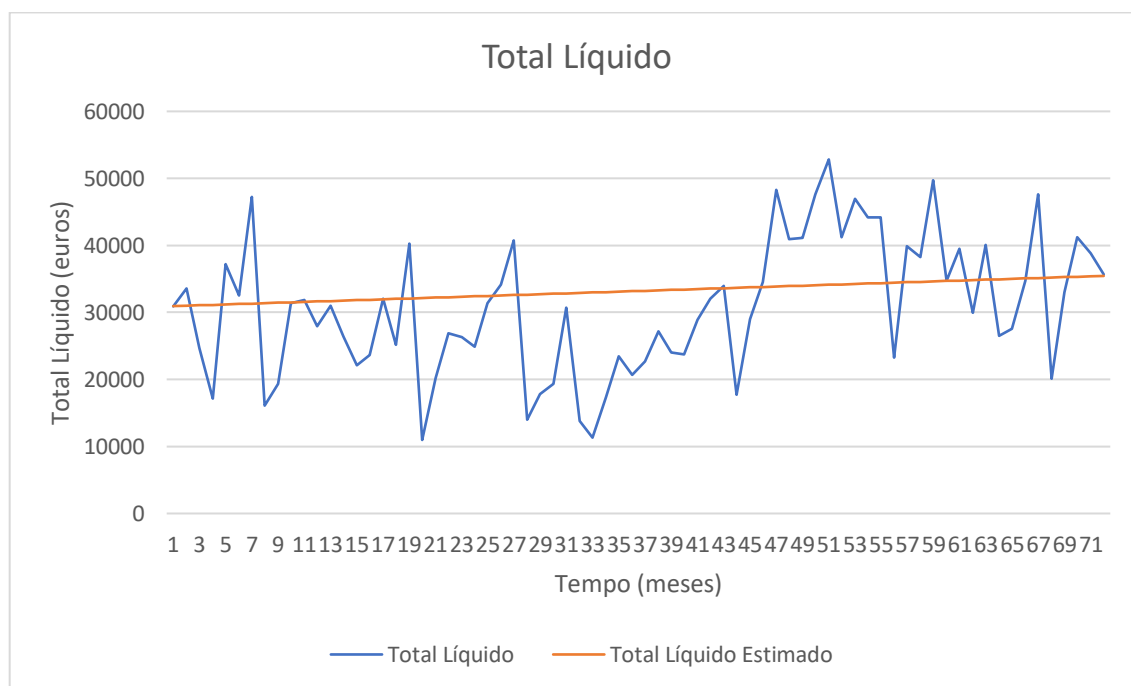


Figura 61 - Representação gráfica da aplicação do método deriva à estimativa do total líquido

6.3.1.5. Método regressivo – modelo 1

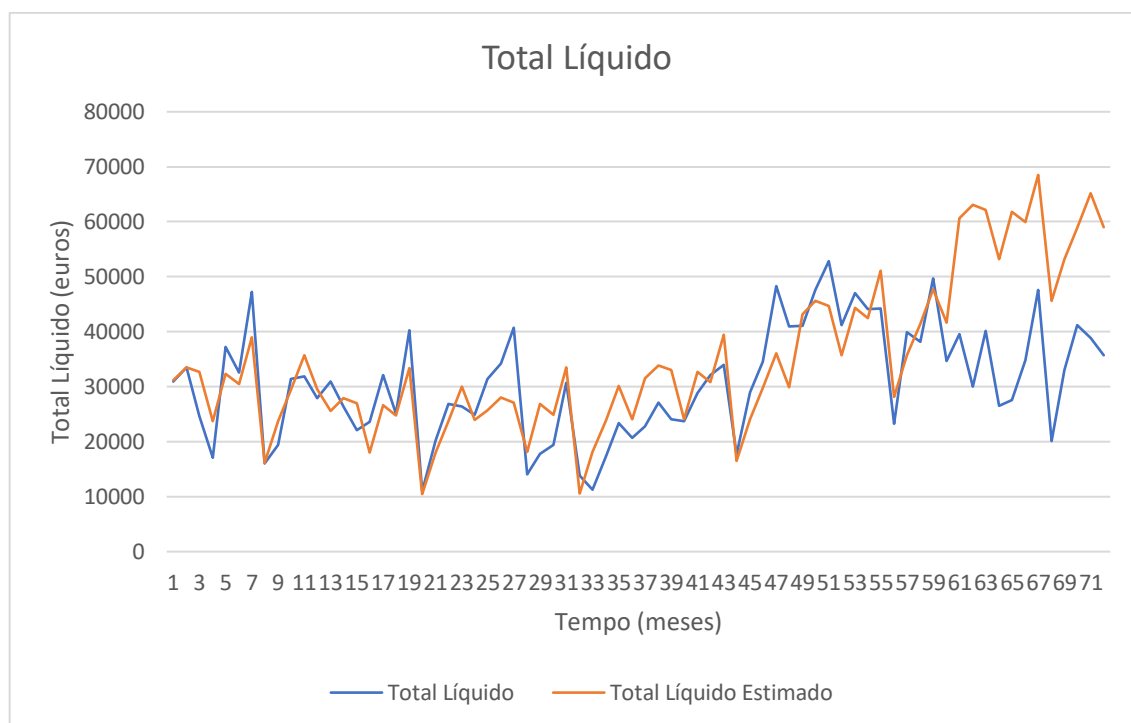


Figura 62 - Representação gráfica da aplicação do modelo 1 à estimativa do total líquido

6.3.1.6. Método regressivo – modelo 2

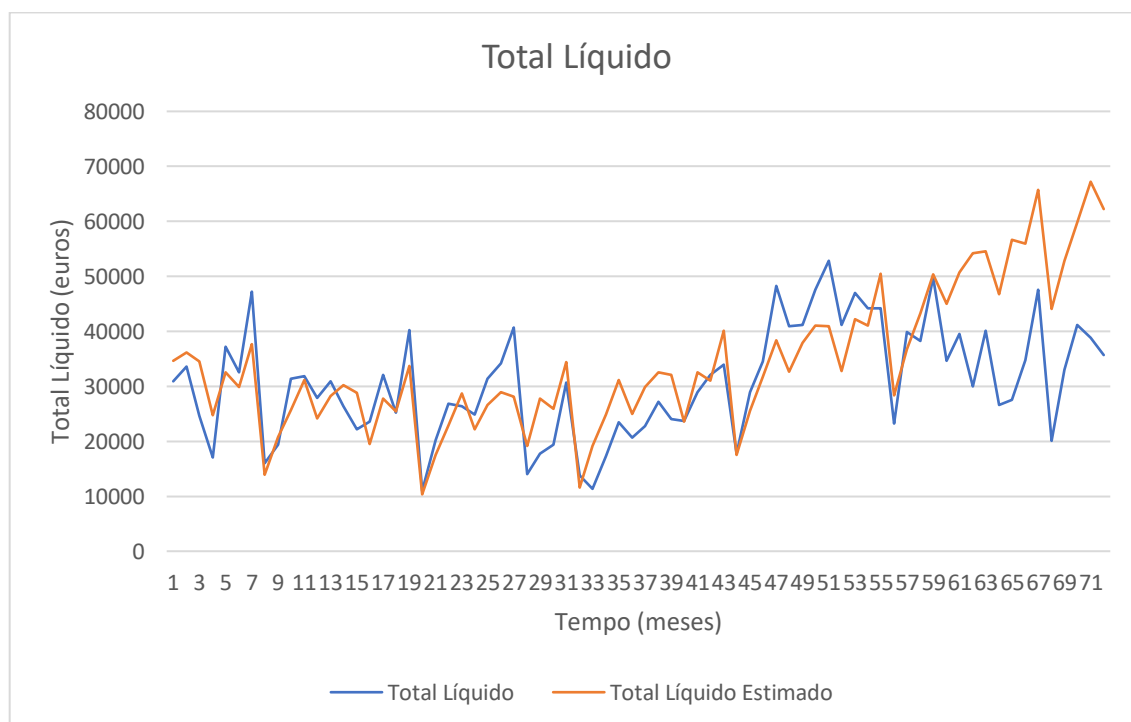


Figura 63 - Representação gráfica da aplicação do modelo 2 à estimativa do total líquido

6.3.1.7. Método regressivo – modelo 3

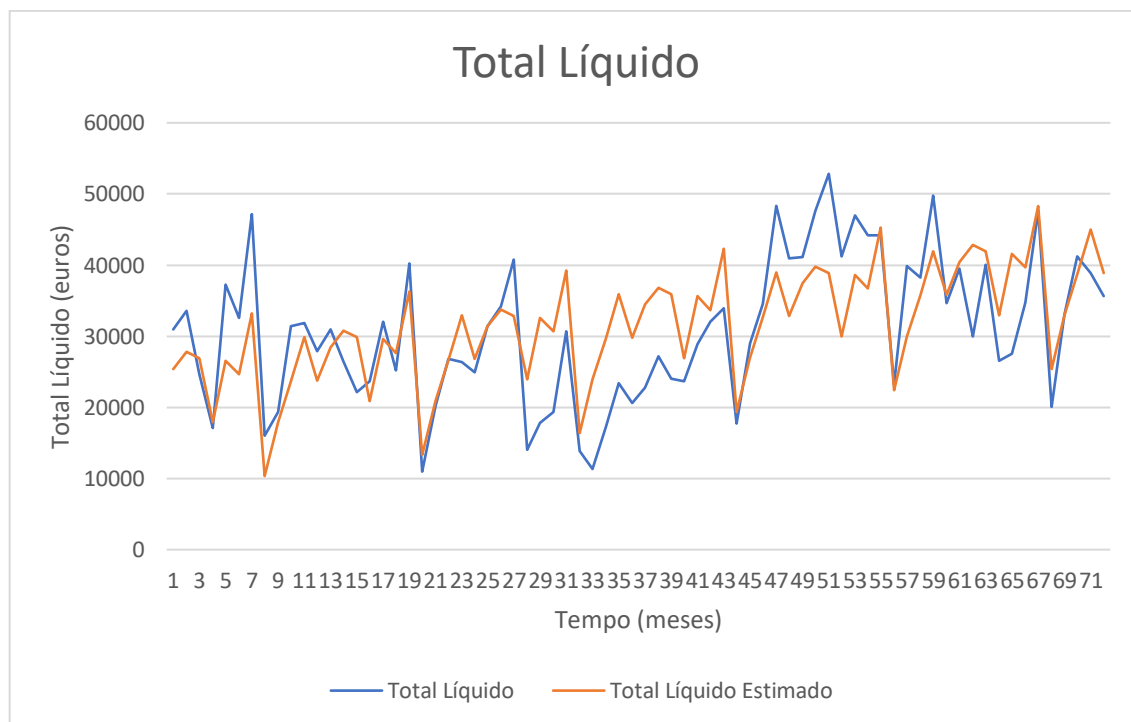


Figura 64 - Representação gráfica da aplicação do modelo 3 à estimativa do total líquido

6.3.1.8. Método regressivo – modelo 4

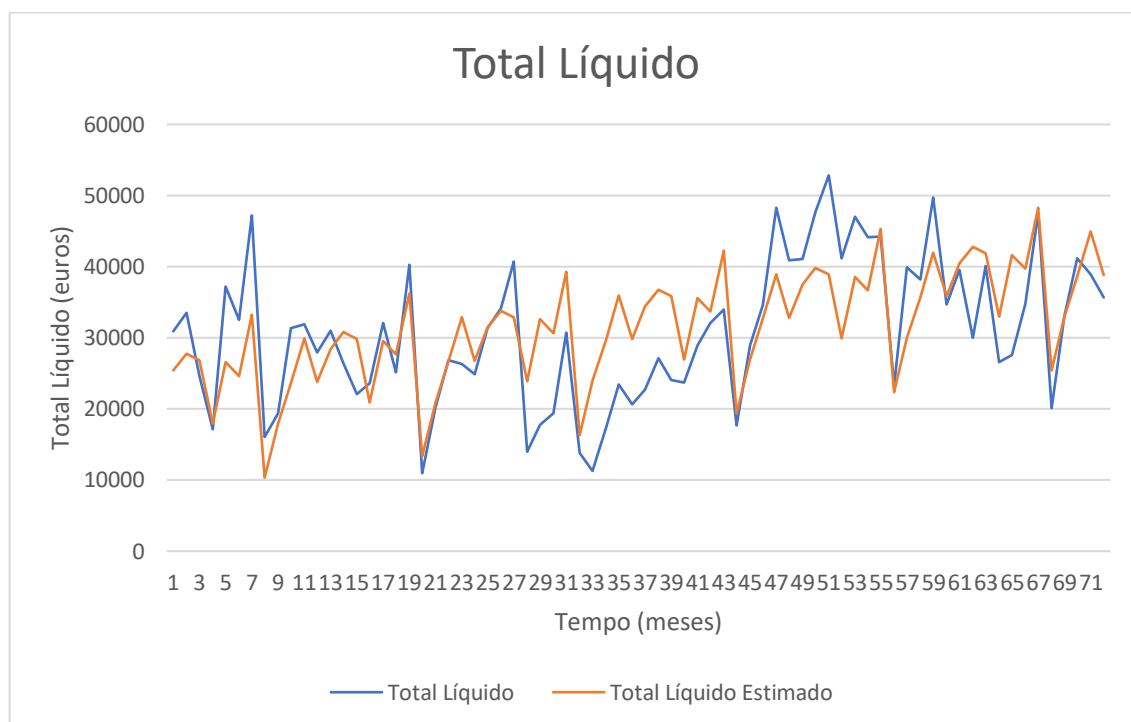


Figura 65 - Representação gráfica da aplicação do modelo 4 à estimativa do total líquido

6.3.1.9. ARIMA

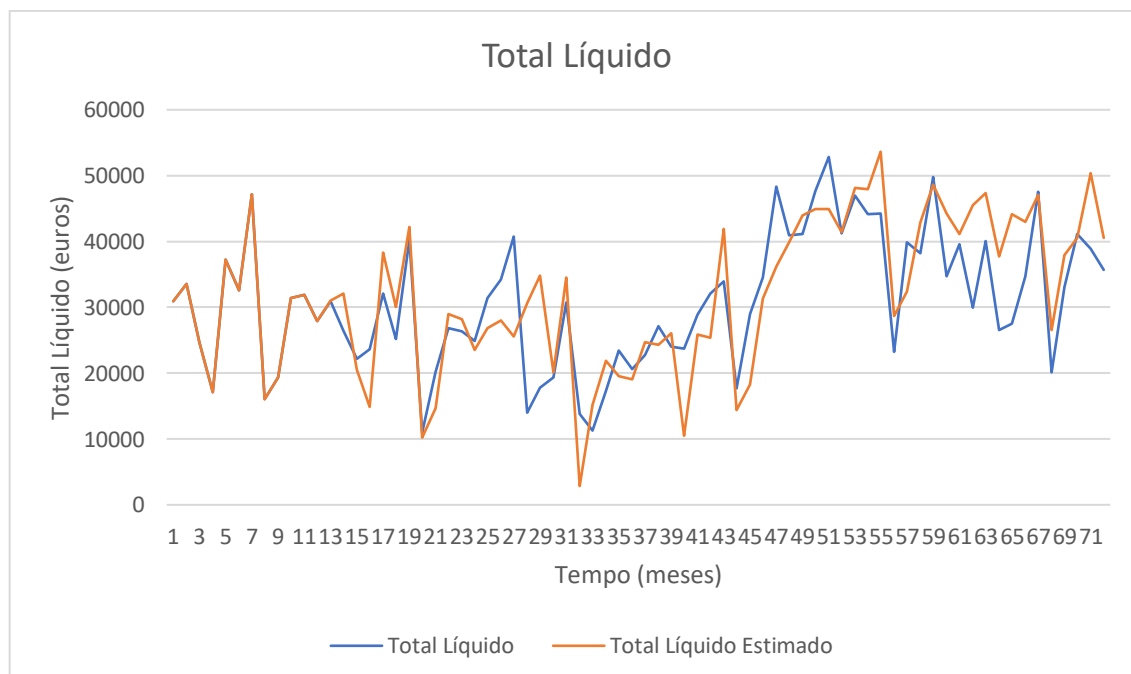


Figura 66 - Representação gráfica da aplicação do método ARIMA à estimativa do total líquido

6.3.1.10. Precisão dos modelos

Após a aplicação de cada um dos modelos quer aos dados de ajuste (período 2018 a 2022) quer aos dados de previsão (2023), foram calculados os erros percentuais correspondentes à precisão do modelo. Os dados obtidos são os seguintes:

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	8 492,82 €	35%	7 198,04 €	20%
Naive	7 773,06 €	34%	5 920,33 €	19%
SNaive	9 493,61 €	32%	8 631,36 €	27%
Drift	8 448,94 €	38%	5 858,69 €	19%
1	4 595,46 €	17%	24 690,60 €	77%
2	4 963,50 €	18%	21 308,96 €	66%
3	5 997,56 €	23%	4 895,21 €	16%
4	5 997,56 €	23%	4 895,21 €	16%
ARIMA	4 239,95 €	17%	7 444,35 €	24%

Tabela 21 - Resumo dos erros de previsão de cada método/modelo para o total líquido global

Os resultados indicam que o modelo 1 e o ARIMA tiveram o melhor desempenho no ajuste histórico (17%), apresentando erros de previsão de 77% e 24%, respetivamente. Os modelos 3 e 4 mostraram os menores erros previstos para 2023 (16%), apesar de terem erros ligeiramente mais elevados no ajuste histórico comparativamente com o modelo ARIMA. Modelos como Naive e Drift tiveram erros maiores na previsão futura e os modelos 1 e 2, embora ajustem bem aos dados de treino, falham nas previsões com erros percentuais altos para 2023 (77% e 66%, respetivamente).

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.1) indicaram que os quatro modelos têm um F estatisticamente significativo sendo, portanto, adequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.1). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.1) permitiu identificar que, neste caso, o modelo Drift produz previsões enviesadas.

6.3.2. Total líquido por cliente

Os resultados obtidos no *clustering* definiram dois *clusters*, sendo que o *cluster 1* compreende 92 casos, representando 96,84% do número de clientes, sendo que o *cluster 2*, compreende 3 casos, correspondendo a 3,16%.

Assim, a análise preditiva foi realizada para os centros dos dois *clusters*, tendo os modelos sido aplicados a ambos. Além das tabelas mostradas a seguir, é possível visualizar as representações gráficas dos resultados na secção 9.2.

6.3.2.1. Cluster 1

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	56,20 €	43%	43,77 €	29%
Naive	53,44 €	41%	58,59 €	33%
SNaive	73,68 €	49%	60,56 €	40%
Drift	63,80 €	51%	65,88 €	35%
1	35,89 €	25%	119,54 €	80%
2	37,50 €	26%	98,46 €	65%
3	42,37 €	31%	38,02 €	26%
4	42,37 €	31%	38,02 €	26%
ARIMA	46,75 €	35%	42,78 €	30%

Tabela 22 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 1 do total líquido por cliente

Os modelos 1 e 2 apresentam os menores erros de ajuste (25% e 26%, respetivamente), mas têm erros de previsão elevados para 2023 (80% e 65%). Já os modelos 3 e 4 mostram um desempenho equilibrado, com erros moderados tanto no ajuste quanto na previsão (31% de erro no ajuste e 26% na previsão), sendo os mais consistentes para 2023. O modelo ARIMA também apresenta uma boa performance com um erro de ajuste de 35% e um erro de previsão de 30%. Por outro lado, modelos como Naive, SNaive, e Drift têm erros de previsão mais elevados, especialmente o SNaive com 40% de erro na previsão para 2023.

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.2) indicaram que os quatro modelos têm um F estatisticamente significativo sendo, portanto, adequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.2). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.2) permitiu identificar que nenhum dos modelos produz previsões enviesadas.

6.3.2.2. Cluster 2

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	1 589,80 €	39%	1 508,37 €	23%
Naive	1 363,04 €	36%	1 615,78 €	34%
SNaive	1 398,39 €	26%	1 919,02 €	37%
Drift	1 313,69 €	36%	1 849,22 €	39%
1	727,29 €	17%	4 564,43 €	85%
2	747,96 €	18%	4 083,47 €	77%
3	983,75 €	23%	1 348,41 €	27%
4	983,75 €	23%	1 348,41 €	27%
ARIMA	848,87 €	20%	3 029,83 €	59%

Tabela 23 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 2 do total líquido por cliente

Os modelos 1 e 2 destacam-se por terem os menores erros de ajuste (17% e 18%, respetivamente), mas apresentam os maiores erros de previsão para 2023 (85% e 77%). O modelo Mean tem desempenho estável, com um erro de ajuste de 39% e de previsão de 23%. Modelos como Naive, SNaive, e Drift têm erros de ajuste e previsão mais elevados, especialmente o Drift, que atinge 39% na previsão. O ARIMA tem um bom ajuste (20%), mas a previsão para 2023 apresenta um erro relativamente alto (59%). Os modelos 3 e 4 mostram um equilíbrio, com erros moderados tanto no ajuste quanto na previsão (23% no ajuste e 27% na previsão).

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.3) indicaram que os quatro modelos têm um F estatisticamente significativo sendo, portanto, adequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.3). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.3) permitiu identificar que os modelos SNaive e Drift produzem previsões enviesadas.

6.3.3. Total líquido por tipo de lavagem

Neste caso, o *clustering* obteve 8 *clusters*, sendo que o *cluster* 1 compreende 13 casos, representando 65,00 % dos 20 tipos de lavagem, no entanto representando apenas menos de 4% do total líquido, e todos os restantes sete *clusters* compreendem 1 caso cada, correspondendo a 5,00 % do total de tipos de lavagem. Além das tabelas mostradas a seguir, é possível visualizar as representações gráficas dos resultados na secção 9.2.

6.3.3.1. Cluster 1

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	57,22 €	135%	62,44 €	10439%
Naive	69,98 €	158%	41,39 €	892%
SNaive	71,96 €	167%	58,97 €	27688%
Drift	76,80 €	73%	42,11 €	724%
1	52,46 €	121%	46,29 €	2031%
2	52,76 €	120%	51,00 €	2258%
3	53,16 €	126%	58,04 €	7854%
4	53,16 €	126%	58,04 €	7854%
ARIMA	57,22 €	135%	62,44 €	10439%

Tabela 24 - Resumo dos erros percentuais de previsão de cada método/modelo para o *cluster* 1 do total líquido por tipo de lavagem

Modelos como Mean e ARIMA têm erros semelhantes, com 135% de erro no ajuste e 10.439% de erro na previsão, sugerindo uma grande divergência nas previsões para 2023. O SNaive também apresenta um erro de previsão extremamente alto (27.688%), seguido pelos modelos 3 e 4, ambos com 7.854%. Os modelos Naive, 1, e 2 apresentam erros menores em comparação, mas ainda muito elevados (892%, 2.031%, e 2.258%, respetivamente). O Drift é o modelo com o menor erro de ajuste (73%) e de previsão (724%), sendo o mais equilibrado entre os demais, embora ainda com erros elevados.

No caso deste *cluster*, todos os modelos apresentam erros percentualmente muito elevados, mas que em termos líquidos corresponde a poucas dezenas de euros. Este facto poderá estar relacionado com o comportamento do total líquido neste *cluster*, que varia entre um máximo de cerca de 300 euros e um mínimo de 0,07 €. É de notar que o erro percentual dá mais importância à estimação de valores mais baixos. Assim, mesmo que o desvio absoluto da estimação seja baixo, o erro percentual pode ser alto. Por exemplo, um desvio absoluto de 0,07 € numa estimativa de 0,07 €, corresponde a um erro percentual de 100%.

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos.

Os resultados (apresentados em 9.3.4) indicaram que os quatro modelos têm um F estatisticamente não significativo sendo, portanto, desadequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.4). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.4) permitiu identificar que o modelo Drift produz previsões enviesadas.

6.3.3.2. Cluster 2 – Tipo de lavagem NORMAL

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	5 057,40 €	88%	3 377,50 €	33%
Naive	2 903,47 €	54%	8 441,29 €	100%
SNaive	4 172,41 €	36%	7 936,49 €	85%
Drift	4 274,94 €	89%	9 780,08 €	114%
1	1 554,71 €	20%	17 115,48 €	184%
2	1 712,71 €	23%	15 506,81 €	168%
3	2 343,46 €	35%	9 235,53 €	104%
4	2 343,46 €	35%	9 235,53 €	104%
ARIMA	2 424,26 €	38%	9 440,46 €	105%

Tabela 25 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 2 do total líquido por tipo de lavagem

O modelo Mean tem um erro de ajuste elevado (88%) mas um erro de previsão menor (33%). Por contraste, o Naive apresenta um erro de ajuste mais baixo (54%) mas um erro de previsão muito alto (100%). O SNaive destaca-se por um bom ajuste (36%) mas com um erro elevado na previsão (85%). Modelos como o Drift, 3, 4, e ARIMA mostram altos erros de previsão acima de 100%. Os modelos 1 e 2 têm os melhores ajustes (20% e 23%, respetivamente), mas falham na previsão para 2023, com erros extremamente altos (184% e 168%). O desempenho na previsão é melhor no Mean, apesar de outros modelos, como SNaive e Naive, se ajustarem melhor aos dados históricos.

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.5) indicaram que os quatro modelos têm um F estatisticamente significativo sendo, portanto, adequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em

9.4.5). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.5) permitiu identificar que os modelos SNaive e Drift produzem previsões enviesadas.

6.3.3.3. Cluster 3 – Tipo de lavagem ESPECIAL 60

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	2 500,06 €	53%	5 563,45 €	43%
Naive	2 342,42 €	44%	5 012,94 €	38%
SNaive	2 861,89 €	47%	3 141,65 €	24%
Drift	2 228,77 €	36%	4 718,15 €	36%
1	1 688,89 €	34%	2 241,50 €	22%
2	1 801,90 €	37%	2 268,40 €	20%
3	1 834,24 €	37%	3 071,02 €	23%
4	1 834,24 €	37%	3 071,02 €	23%
ARIMA	1 740,20 €	32%	3 007,48 €	23%

Tabela 26 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 3 do total líquido por tipo de lavagem

O ARIMA é o modelo com o melhor desempenho geral, com o menor erro de ajuste (32%) e um erro de previsão de 23%, próximo de outros modelos. Modelos como 1 e 2 também se destacam com erros baixos quer no ajuste (34% e 37%) quer na previsão (22% e 20%, respetivamente), demonstrando boa consistência. O SNaive tem um ajuste razoável (47%) e o menor erro de previsão (24%), enquanto os modelos Mean, Naive e Drift apresentam erros de ajuste moderados, entre 36% e 53%, mas previsões menos precisas, com erros de previsão a variar entre 36% e 43%. Os modelos 3 e 4 têm desempenho semelhante ao ARIMA, mas com um erro de ajuste ligeiramente superior.

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.6) indicaram que os quatro modelos têm um F estatisticamente significativo sendo, portanto, adequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.6). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.6) permitiu identificar que os modelos SNaive e Drift produzem previsões enviesadas.

6.3.3.4. Cluster 4 – Tipo de lavagem SILICONE

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	1 912,79 €	55%	1 867,39 €	57%
Naive	1 688,05 €	50%	1 681,98 €	45%
SNaive	2 064,47 €	46%	3 974,43 €	134%
Drift	1 877,63 €	49%	1 689,68 €	45%
1	1 235,66 €	32%	7 154,70 €	231%
2	1 268,55 €	31%	5 704,96 €	184%
3	1 453,55 €	36%	3 103,82 €	100%
4	1 453,55 €	36%	3 103,82 €	100%
ARIMA	1 394,01 €	40%	2 609,46 €	81%

Tabela 27 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 4 do total líquido por tipo de lavagem

Os modelos Mean, Naive e Drift têm erros de ajuste e previsão semelhantes, variando entre 45% e 57%. O SNaive tem um ajuste razoável (46%), mas a previsão é imprecisa, com um erro elevado de 134%. Os modelos 1 e 2 apresentam os melhores ajustes (32% e 31%), mas falham significativamente na previsão de 2023, com erros elevados (231% e 184%, respetivamente). O ARIMA apresenta um erro de ajuste mais baixo (40%) e um erro de previsão um pouco elevado (81%), enquanto os modelos 3 e 4 têm desempenho moderados no ajuste (36%), mas com um erro elevado na previsão (100%).

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.7) indicaram que os quatro modelos têm um F estatisticamente significativo sendo, portanto, adequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.7). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.7) permitiu identificar que nenhum dos modelos produz previsões enviesadas.

6.3.3.5. Cluster 5 – Tipo de lavagem SECO

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	1 446,64 €	121%	2 106,86 €	106%
Naive	1 592,02 €	87%	2 529,78 €	64%
SNaive	2 094,43 €	156%	1 604,27 €	50%
Drift	1 608,53 €	64%	2 558,12 €	65%
1	1 277,41 €	95%	2 033,04 €	67%
2	1 275,41 €	93%	1 913,77 €	70%
3	1 277,78 €	95%	2 059,07 €	65%
4	1 277,78 €	95%	2 059,07 €	65%
ARIMA	1 397,70 €	108%	2 069,98 €	97%

Tabela 28 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 5 do total líquido por tipo de lavagem

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.8) indicaram que os quatro modelos têm um F estatisticamente não significativo sendo, portanto, desadequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.8). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.8) permitiu identificar que o modelo Drift produz previsões enviesadas.

6.3.3.6. Cluster 6 – Tipo de lavagem MICRONIZAR

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	1 193,55 €	204%	1 208,10 €	191%
Naive	1 028,03 €	105%	934,27 €	145%
SNaive	1 300,40 €	127%	782,50 €	74%
Drift	1 292,59 €	258%	691,21 €	103%
1	822,03 €	75%	1 322,18 €	182%
2	816,23 €	72%	969,93 €	125%
3	814,30 €	72%	940,00 €	120%
4	814,30 €	72%	940,00 €	120%
ARIMA	1 028,10 €	104%	934,27 €	145%

Tabela 29 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 6 do total líquido por tipo de lavagem

O modelo SNaive destaca-se com o menor erro de previsão, ainda assim elevado (74%) e um erro de ajuste elevado (127%). O modelo Drift, apesar de ter o maior erro de ajuste (258%), apresenta um erro de previsão consideravelmente mais baixo (103%). Os modelos 1, 2, 3 e 4 têm erros de ajuste semelhantes, em torno de 72%-75%, mas os erros de previsão variam entre 120% e 182%. Os modelos Naive e ARIMA apresentam erros de ajuste próximos (104%-105%) e erros de previsão de 145%, enquanto o Mean tem os maiores erros tanto no ajuste (204%) e na previsão (191%).

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.9) indicaram que os quatro modelos têm um F estatisticamente significativo sendo, portanto, adequados.

No caso do modelo ARIMA, o ajuste automático feito pelo software conduziu a um modelo ARIMA (0,1,0), ou seja, a um passeio aleatório e, por este motivo, não se aplicou o teste de Box-Ljung.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.9) permitiu identificar que o modelo Drift produz previsões enviesadas.

6.3.3.7. Cluster 7 – Tipo de lavagem ANTIPIILLING

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	2 029,06 €	1923%	1 053,27 €	607%
Naive	1 346,45 €	511%	722,26 €	91%
SNaive	1 898,48 €	2878%	550,79 €	173%
Drift	3 498,77 €	5728%	1 783,60 €	425%
1	1 480,58 €	1192%	2 892,28 €	1210%
2	1 276,65 €	1112%	3 070,04 €	1108%
3	1 763,98 €	1862%	2 497,70 €	926%
4	1 763,98 €	1862%	2 497,70 €	926%
ARIMA	1 302,46 €	821%	566,54 €	117%

Tabela 30 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 7 do total líquido por tipo de lavagem

A análise da tabela de erros do modelo de previsão revela erros percentuais extremamente elevados quer no período de ajuste (2018-2022) quer na previsão para 2023. No ajuste, todos os modelos apresentam erros muito elevados, com destaque para o modelo Drift, com 5728%, e o SNaive, com 2878%. A previsão para 2023 também apresenta erros elevados em todos os modelos, embora menores em comparação ao ajuste. O modelo Naive apresenta o menor erro percentual (91%) em 2023, seguido pelo ARIMA (117%) e o SNaive (173%). Apesar dos modelos apresentarem erros percentuais elevados, em termos de valor

líquido correspondente, o erro anda na ordem dos poucos milhares de euros (variando entre cerca de 500 € e 3500 €).

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.10) indicaram que os quatro modelos têm um F estatisticamente significativo sendo, portanto, adequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.10). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.10) permitiu identificar que os modelos SNaive e Drift produzem previsões enviesadas.

6.3.3.8. Cluster 8 – Tipo de lavagem C/DISPERSANTE

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	691,37 €	128%	510,84 €	79%
Naive	846,25 €	119%	524,27 €	85%
SNaive	1 125,44 €	133%	664,94 €	80%
Drift	668,79 €	118%	525,63 €	86%
1	659,20 €	108%	645,17 €	62%
2	658,16 €	109%	675,00 €	74%
3	655,24 €	111%	632,84 €	91%
4	655,24 €	111%	632,84 €	91%
ARIMA	644,80 €	121%	539,12 €	84%

Tabela 31 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 8 do total líquido por tipo de lavagem

Os dados revelam que no período de ajuste (2018-2022), os erros variam entre 108% e 133%, considerando-se erros muito elevados. Para a previsão de 2023, os erros são menores, mas ainda assim elevados, com todos os modelos a apresentarem valores inferiores a 91%, sendo que o modelo 1 apresenta o erro percentual mais baixo (62%). Embora os erros sejam consideravelmente menores na previsão, ainda são elevados. No entanto, mais uma vez, os erros apesar de percentualmente elevados, correspondem a poucas centenas de euros em termos líquidos.

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.11) indicaram que os quatro modelos têm um F estatisticamente não significativo sendo, portanto, desadequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.11). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.11) permitiu identificar que nenhum dos modelos produz previsões enviesadas.

6.3.4. Total líquido por tipo de peça

Neste caso, o *clustering* obteve 2 *clusters*, em que o *cluster* 1 compreende 42 casos, representando 96,67 % do total de peças, sendo que o *cluster* 2, compreende 1 caso, o tipo de peça T-SHIRT, correspondendo a 2,33 % do total de peças. No entanto, em termos de valor do total líquido, o *cluster* 2 representa mais de 48 % do total. Além das tabelas mostradas a seguir, é possível visualizar as representações gráficas dos resultados na secção 9.2.

6.3.4.1. Cluster 1

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	113,14 €	39%	99,51 €	21%
Naive	108,68 €	39%	95,70 €	22%
SNaive	154,12 €	43%	127,66 €	32%
Drift	107,80 €	33%	94,83 €	23%
1	75,33 €	25%	180,59 €	49%
2	76,27 €	25%	148,72 €	40%
3	77,03 €	25%	98,41 €	26%
4	77,03 €	25%	98,41 €	26%
ARIMA	99,97 €	33%	95,92 €	25%

Tabela 32 - Resumo dos erros percentuais de previsão de cada método/modelo para o *cluster* 1 do total líquido por tipo de peça

No período de ajuste (2018-2022), quase todos os modelos apresentam erros em torno de 25% a 43%, com destaque para os modelos 1, 2, 3 e 4, cujos erros estão abaixo de 30%. Para a previsão de 2023, os erros percentuais são menores, variando de 21% a 49%, com modelos como Mean, Naive, Drift, 3, 4 e ARIMA a apresentarem erros inferiores a 30%. Apenas os modelos 1, 2 e SNaive ultrapassam ligeiramente este valor, com erros mais elevados de 32% a 49%.

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos.

Os resultados (apresentados em 9.3.12) indicaram que os quatro modelos têm um F estatisticamente significativo sendo, portanto, adequados.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.12). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.12) permitiu identificar que o modelo Drift produz previsões enviesadas.

6.3.4.2. Cluster 2 – Tipo de peça T-SHIRT

Modelo	Erro			
	Ajuste (2018-2022)		Previsão (2023)	
	MAE	MAPE	MAE	MAPE
Mean	5 002,31 €	45%	4 951,88 €	39%
Naive	4 546,92 €	41%	3 946,61 €	40%
SNaive	5 388,67 €	40%	6 421,99 €	53%
Drift	6 924,35 €	72%	4 047,95 €	40%
1	2 600,53 €	19%	17 379,68 €	134%
2	2 783,83 €	20%	15 489,23 €	122%
3	4 202,69 €	33%	2 776,44 €	26%
4	4 202,69 €	33%	2 776,44 €	26%
ARIMA	3 198,82 €	24%	5 960,60 €	54%

Tabela 33 - Resumo dos erros percentuais de previsão de cada método/modelo para o cluster 2 do total líquido por tipo de peça

A tabela mostra uma variação significativa nos erros percentuais dos modelos tanto no ajuste (2018-2022) quanto na previsão para 2023. No período de ajuste, a maioria dos modelos apresenta erros entre 19% e 45%, com exceção do Drift (72%), que tem um erro elevado. O modelo 1 destaca-se com o menor erro percentual no ajuste (19%), enquanto outros, como o Mean, Naive e SNaive, mostram erros moderados na faixa de 40% a 45%. Para a previsão de 2023, observa-se um aumento no erro de alguns modelos, especialmente o 1 e o 2, que apresentam erros elevados de 134% e 122%, respetivamente. Outros modelos, como o Mean, Naive, Drift, 3 e 4, mantêm erros moderados entre 26% e 40%. Os modelos ARIMA e o SNaive apresentam erros elevados, com 54% e 53%, respetivamente. Em geral, a performance dos modelos na previsão de 2023 é mais heterogénea, com alguns bons resultados e outros com erros consideráveis.

Após a aplicação de quatro modelos regressivos à série temporal em análise, avaliou-se a significância estatística tanto dos coeficientes individuais como dos modelos completos. Os resultados (apresentados em 9.3.13) indicaram que os modelos 1 e 2 têm um F estatisticamente significativo sendo, portanto, adequados. Os modelos 3 e 4, tendo um F

estatisticamente não significativo, poderiam ser considerados desadequados. No entanto, tendo em conta que estes modelos têm um valor de p associado ao coeficiente F de 0,0520, ou seja, apenas duas milésimas superior ao limite para serem estatisticamente significativos, não se considera como tal.

A adequabilidade do modelo ARIMA foi avaliada com base no teste de Box-Ljung, resultando um valor de qui-quadrado com $p > 0,05$ (conforme resultados apresentados em 9.4.13). Assim, conclui-se que o modelo ARIMA é estatisticamente adequado para descrever os dados.

A análise sumária dos resíduos dos modelos de previsão (apresentados em 9.5.13) permitiu identificar que o modelo Drift produz previsões enviesadas.

6.3.5. Análise global dos resultados

Nas secções anteriores foram analisados individualmente todos os *clusters*, sendo possível verificar qual o modelo de previsão que fornecia resultados com erros menores e qual o valor desses erros. A tabela seguinte resume a informação relativa à análise de todos os casos, apresentando os melhores modelos de previsão para o total líquido global e para cada um dos *clusters*. Quer no ajuste, quer na previsão foram seleccionados os modelos com menor erro percentual (MAPE). É de notar que, na maior parte das vezes, o modelo com melhor ajuste não é o que tem melhor previsão.

Dados/Cluster	N.º clientes/ lavagens /peças	% do TL Global	Ajuste			Previsão		
			Modelo	MAE	MAPE	Modelo	MAE	MAPE
TL Global	-	100%	ARIMA	4 239,95 €	17%	3	4 895,21 €	16%
TL/Cliente C1	92	53%	1	35,89 €	25%	3	38,02 €	26%
TL/Cliente C2	3	47%	1	727,29 €	17%	Mean	1 508,37 €	23%
TL/Lavagem C1	13	4%	Drift	76,80 €	73%	Drift	42,11 €	724%
TL/Lavagem C2	1	34%	1	1 554,71 €	20%	Mean	3 377,50 €	33%
TL/Lavagem C3	1	21%	ARIMA	1 740,20 €	32%	2	2 268,40 €	20%
TL/Lavagem C4	1	16%	2	1 268,55 €	31%	Drift	1 689,68 €	45%
TL/Lavagem C5	1	9%	Drift	1 608,53 €	64%	SNaive	1 604,27 €	50%
TL/Lavagem C6	1	7%	3	814,30 €	72%	SNaive	782,50 €	74%
TL/Lavagem C7	1	6%	Naive	1 346,45 €	511%	Naive	722,26 €	91%
TL/Lavagem C8	1	4%	1	659,20 €	108%	1	645,17 €	62%
TL/Peça C1	42	52%	1	75,33 €	25%	Mean	99,51 €	21%
TL/Peça C2	1	48%	2	2 600,53 €	19%	3	2 776,44 €	26%

Tabela 34 - Análise global dos resultados dos modelos de previsão aplicados ao total líquido global e a cada cluster

6.3.5.1. Análise do MAE e do MAPE

O MAE permite uma análise direta do erro absoluto em unidades monetárias, o que facilita a interpretação dos desvios das previsões em relação aos valores reais. O MAPE mede o erro percentual médio, que é útil para avaliar a precisão do modelo em termos relativos, ajustando-se às escalas dos dados.

- **Total líquido global:** para o total líquido global os modelos que melhores resultados fornecem são para o ajuste o modelo ARIMA (4239,95 €, 17%) e o modelo 3, para a previsão (4895,21 €, 16%).
- **Total líquido por cliente:** no caso do *cluster* 1, que representa 92 clientes e 53% do total líquido, verifica-se que o modelo que apresenta melhores resultados no ajuste é o modelo 1 (35,89 €, 25%) e na previsão o modelo 3 (38,02 €, 26%). No caso do *cluster* 2, que representa apenas 3 clientes, mas 47 % do total líquido, o modelo 1 apresenta melhores resultados no ajuste (727,29 €, 17%) e o Mean melhores resultados na previsão (1508,37 €, 23%).
- **Total líquido por tipo de lavagem:** no *cluster* 1, que representa 13 tipos de lavagem, mas apenas 4% do total líquido, o modelo com melhores resultados é o Drift, com erros de 76,80 € (73%) no ajuste e 42,11 € (724%) na previsão. O *cluster* 2, representa o tipo de lavagem NORMAL, responsável por 34 % do total líquido, tendo como melhor modelo no ajuste o modelo 1 (1554,71 €, 20%) e na previsão o Mean (3377,50 €, 33%). Para o caso do *cluster* 3 o modelo ARIMA apresenta os melhores resultados no ajuste (1740,20 €, 32%) e na previsão o modelo 2 (2268,40 €, 20%). No *cluster* 4, o modelo com melhores resultados no ajuste é o 2 (1268,55 €, 31%) e na previsão o modelo Drift (1689,68 €, 45%). No caso do *cluster* 5, o melhor modelo no ajuste é o Drift (1608,53 €, 64%) e na previsão o modelo SNaive (1604,27 €, 50%). Para o *cluster* 6, os melhores resultados no ajuste são obtidos pela aplicação do modelo 3 (814,30 €, 72%) e na previsão pelo modelo SNaive (782,50 €, 74%). No *cluster* 7, o modelo Naive apresenta os melhores resultados quer no ajuste quer na previsão (1346,45 €, 511% e 722,26 €, 91%, respetivamente). Por fim, para o *cluster* 8, o modelo que apresenta melhores resultados, no ajuste e na previsão, é o modelo 1 (659,20 €, 108% e 645,17 €, 62%, respetivamente).
- **Total líquido por tipo de peça:** no *cluster* 1, que representa 42 tipos de peça, o modelo que produz melhores resultados no ajuste é o modelo 1 (75,33 €, 25%) e na previsão o modelo Mean (95,51 €, 21%). No caso do *cluster* 2, que integra apenas o tipo de peça T-SHIRT, o modelo com melhores resultados no ajuste é o modelo 2 (2600,53 €, 19%) e na previsão o modelo 3 (2776,44 €, 26%).

A análise do MAE e do MAPE revela variações consideráveis entre a precisão dos modelos para cada um dos *clusters*. Na grande maioria dos casos, o modelo que produz melhores resultados no ajuste, não é o modelo que garante os melhores resultados quando aplicado à previsão.

Quando associados os erros de cada modelo, quer no ajuste quer na previsão, à percentagem do total líquido representado por aquele *cluster*, verifica-se que nos casos em que a percentagem do total líquido é mais elevada a precisão do modelo é melhor. Os casos em que os erros dos modelos de previsão são mais elevados correspondem aos *clusters* 1, 5, 6, 7 e 8 do total líquido por tipo de lavagem, que representam percentagens baixas do total líquido, todas inferiores a 10%.

De uma forma geral verifica-se que os modelos regressivos e autorregressivos produzem melhores resultados no ajuste, mas que ao nível da previsão existe divisão entre os modelos regressivos e os métodos simples.

7. Conclusões

O conceito de empresa orientada por dados refere-se a organizações que integram dados nos seus processos de decisão, utilizando tecnologias avançadas para identificar oportunidades e melhorar a eficiência. Embora muitas empresas recolham grandes volumes de dados, poucas conseguem extrair valor significativo devido à falta de análise e processamento adequados. As empresas com uma cultura orientada por dados destacam-se pela capacidade de tomar decisões mais informadas, adaptarem-se rapidamente às mudanças e promoverem a inovação, recorrendo a ferramentas como análises preditivas e prescritivas.

Os avanços em sensores, armazenamento e *data mining* têm revolucionado a forma como dados complexos são gerados e analisados, tornando impraticável a análise manual. Esta evolução impulsionou o desenvolvimento de soluções automatizadas, como algoritmos de *machine learning* e técnicas de *clustering*, essenciais para a gestão de *big data*. Estas ferramentas permitem identificar padrões e tendências que orientam a tomada de decisões estratégicas.

A capacidade de extrair informações significativas a partir de *big data* é fundamental em áreas como negócios, ciência, medicina e administração pública, onde a qualidade das decisões depende diretamente da eficácia da análise de dados. No contexto do *machine learning*, a aprendizagem não supervisionada destaca-se como uma abordagem relevante, sendo o *clustering* amplamente utilizado para identificar padrões desconhecidos e segmentar dados com características semelhantes. Esta técnica encontra aplicações diversificadas, incluindo a análise de mercado, a segmentação de clientes e a deteção de padrões em grandes volumes de dados.

Entre os métodos de *clustering*, destacam-se o hierárquico, que organiza os dados numa estrutura em árvore, e o partitivo, exemplificado pelo algoritmo K-médias. Este último divide os dados em *clusters* com base na proximidade aos centróides. Como alternativa, o algoritmo K-medóides utiliza medóides como pontos de referência.

Importa salientar que não existe um algoritmo universal de *clustering*, sendo crucial validar as partições para determinar a configuração mais adequada. Métodos como o *elbow method* e o coeficiente silhueta são amplamente utilizados para avaliar a qualidade dos agrupamentos. O coeficiente silhueta analisa a separação e a proximidade entre *clusters*, permitindo validar a adequação da segmentação. A escolha cuidadosa de algoritmos e métricas é determinante para garantir resultados precisos e úteis na análise de dados complexos.

A análise de dados organizados em séries temporais beneficia de métricas avançadas como o DTW, que alinha séries temporais de diferentes comprimentos e formas, oferecendo maior robustez comparativamente à distância euclidiana.

A previsão de séries temporais constitui uma técnica essencial na ciência de dados, permitindo às organizações planejar com maior precisão, melhorar a alocação de recursos, estabelecer metas estratégicas e detetar anomalias. Baseada na análise de dados históricos recolhidos, esta metodologia identifica padrões como tendências, sazonalidades e ciclos repetitivos, sendo amplamente utilizada em setores como a indústria, o comércio e a administração pública. A sua aplicação é particularmente vantajosa para empresas que procuram otimizar operações, reduzir custos e reforçar a competitividade, como uma lavanderia industrial.

A análise de séries temporais inicia-se pela identificação de padrões que orientam a escolha do método de previsão mais adequado. Entre os padrões mais comuns incluem-se o padrão horizontal, em que os valores oscilam em torno de uma média constante; o padrão de tendência, que reflete movimentos ascendentes ou descendentes ao longo do tempo; o padrão sazonal, caracterizado por flutuações periódicas regulares; o padrão de tendência sazonal, que combina tendências de longo prazo com variações sazonais; e o padrão cíclico, evidenciando variações em ciclos mais longos frequentemente associados a fatores económicos.

Os métodos de previsão variam em complexidade, desde abordagens simples a técnicas sofisticadas. Entre os métodos simples destacam-se o método da Mean, o método Naive, o método SNaive e o método Drift. Para séries temporais mais complexas, métodos avançados como a análise de regressão e o modelo ARIMA (*Autoregressive Integrated Moving Average*) são mais indicados. A análise de regressão modela a relação entre variáveis preditivas e o resultado esperado, enquanto o ARIMA ajusta séries temporais não estacionárias para prever tendências e sazonalidades.

A avaliação da precisão das previsões é essencial, sendo realizada através de métricas como o *Mean Absolute Error* (MAE) e o *Mean Absolute Percentage Error* (MAPE). Estas métricas quantificam o desvio entre as previsões e os valores reais, permitindo selecionar o modelo mais adequado para a série temporal em análise.

Assim, a previsão de séries temporais proporciona múltiplos benefícios, incluindo o planeamento eficaz de recursos, a gestão otimizada de *stocks*, a adaptação dos processos operacionais às flutuações da procura e a definição de estratégias de *marketing* mais precisas. Além disso, a deteção de anomalias pode prevenir problemas operacionais, enquanto uma maior capacidade de resposta às dinâmicas do mercado reforça a competitividade da empresa. Ao oferecer informação detalhada sobre o comportamento futuro da organização, a previsão de séries temporais contribui decisivamente para a eficiência e sustentabilidade de empresas, assegurando um crescimento contínuo e estratégico.

A Precioso & Morais, Lda., uma microempresa, lavandaria industrial fundada em 1997 em Braga, integra o setor têxtil, que presta serviços de lavagem e acabamento têxtil, como lavagem, secagem, tratamento, dobragem e transporte. Empregando 15 trabalhadores, presta serviços sobretudo a empresas do setor têxtil na região do Cávado.

A empresa tem investido na modernização de equipamentos e na eficiência energética, destacando-se o uso de energias renováveis, como energia solar e biomassa, reduzindo a dependência de combustíveis fósseis. No âmbito social, promove a inclusão laboral, apoia trabalhadores-estudantes na gestão dos seus horários e colabora com causas sociais, como o Banco Alimentar Contra a Fome. Assim, a Precioso & Morais combina inovação operacional, sustentabilidade ambiental e responsabilidade social.

Com o objetivo de transformar a lavandaria industrial numa organização orientada por dados, foi aplicada a técnica de *clustering* para analisar os dados de faturação recolhidos entre 2018 e 2023. Esta abordagem permitiu agrupar os dados em conjuntos com características semelhantes, facilitando a identificação de padrões e tendências nos serviços prestados e no comportamento dos clientes. Através desta análise, foi possível revelar informações relevantes, como as preferências de determinados clientes por tipos específicos de lavagem, associações entre tipos de peças e serviços e a sazonalidade na procura dos serviços. Paralelamente, recorreu-se à previsão de séries temporais para prever o comportamento futuro da empresa e promover uma gestão mais informada e orientada por dados. Esta análise baseou-se na organização dos dados de faturação em séries temporais, considerando a evolução mensal dos valores de 2018 a 2022, com os dados de 2023 utilizados para validar os modelos desenvolvidos. A previsão focou-se no total líquido mensal, variável essencial para compreender o desempenho financeiro da empresa e projetar as receitas futuras.

Este estudo baseia-se em dados de faturação recolhidos ao longo de seis anos (2018-2023), com o objetivo de encontrar padrões relevantes no comportamento dos clientes, tipos de lavagem e peças, a fim de melhorar a tomada de decisões e otimizar o modelo de negócio da empresa.

A primeira etapa envolveu a extração e organização dos dados provenientes das guias de transporte emitidas pela empresa. Estes dados incluíam informações sobre a data de emissão, tipo de lavagem, tipo de peça, quantidade, cliente e o valor total líquido de cada serviço. Com base nessa organização, o estudo focou-se na análise do total líquido em função de três variáveis: cliente, tipo de peça e tipo de lavagem. Os dados foram reorganizados em períodos mensais (totalizando 72 períodos) para que a variação do total líquido pudesse ser analisada como séries temporais.

7.1. Resultados

O objetivo da aplicação dos algoritmos de *clustering* foi encontrar similaridades entre clientes, tipos de peça e tipos de lavagem, em função do total líquido, agrupando-os em *clusters* que permitissem uma posterior análise detalhada. Para tal, foram utilizados dois algoritmos de *clustering*: o k-médias, com a distância euclidiana, e o k-medóides, combinado com a distância temporal obtida pelo *Dynamic Time Warping* (DTW). Além disso, os dados foram analisados com e sem standardização, de modo a comparar os efeitos desse tratamento sobre os resultados.

Os principais resultados obtidos com a aplicação dos algoritmos de *clustering* indicaram que:

- **Total líquido por cliente:** O número ótimo de *clusters* encontrado foi dois, independentemente do algoritmo utilizado. O algoritmo k-médias com distância euclidiana e standardização dos dados apresentou o coeficiente de silhueta mais elevado (0,9429), sugerindo uma melhor separação dos *clusters*. No *cluster* 1 estão incluídos 92 casos (representando 96,84% do total líquido), enquanto no *cluster* 2 existem 3 casos (3,16%).
- **Total líquido por tipo de lavagem:** O número ótimo de *clusters* encontrado foi oito, com o algoritmo k-médias sem standardização apresentando o melhor coeficiente de silhueta (0,9553). O *cluster* 1 compreende 13 dos 20 tipos de lavagem (65% do total), enquanto os restantes sete *clusters* contêm apenas um tipo de lavagem cada (5%).
- **Total líquido por tipo de peça:** O número ótimo de *clusters* encontrado foi dois, com o k-médias sem standardização a obter o coeficiente de silhueta mais elevado (0,9869). O *cluster* 1 contém 42 dos 43 tipos de peça (97,67% dos clientes), enquanto o *cluster* 2 tem apenas um tipo de peça, o tipo T-SHIRT (2,33% dos clientes).

Uma análise detalhada dos resultados obtidos através dos algoritmos de *clustering* aplicados aos dados de faturação, foi realizada com o objetivo de caracterizar os *clusters* e identificar padrões de comportamento nas variáveis relacionadas com clientes, tipos de peças e tipos de lavagem.

Na análise do total líquido por cliente, o algoritmo k-médias com distância euclidiana e dados standardizados revelou a existência de dois *clusters* principais. O *cluster* 1 compreende a maioria dos clientes, com 92 casos que representam 96,84% do total de clientes, enquanto o *cluster* 2 inclui apenas três clientes (34, 176 e 209), correspondendo a 3,16% do número total. A importância destes três clientes, que representam uma parcela tão pequena em termos de número, é destacada pela sua contribuição para o volume de negócios: juntos, eles são responsáveis por 47,56% das receitas totais entre 2018 e 2022.

O cliente 34 é o principal contribuinte para a faturação da empresa, com um volume de negócios de 390.421 euros durante o período analisado, representando 21,67% do total. Este cliente apresentou uma ligeira queda nas receitas entre 2018 e 2020, seguida por uma recuperação significativa em 2021 e 2022. O cliente 176 também foi importante ao nível das receitas, com um crescimento contínuo ao longo dos anos e uma contribuição de 262.011 euros (14,54% do total). O cliente 209, embora com uma trajetória mais estável, gerou 204.360 euros, o que equivale a 11,34% do total de receitas. Estes três clientes destacam-se não apenas pelo volume de receitas, mas também pelo comportamento heterogéneo ao longo do período analisado, com fases de crescimento e decréscimo do volume de negócio distintas.

A análise dos tipos de peças lavadas por cliente revelou que o cliente 34 coloca principalmente POLO (60%) e T-SHIRT (33%) na lavandaria, enquanto o cliente 176 lava maioritariamente T-SHIRT (51%), SWEAT (21%) e FELPA (15%). O cliente 209, por sua vez, foca-se maioritariamente em T-SHIRT (69%), com um segundo tipo menos representado, o POLO (11%). Este padrão de consumo mostra que, apesar da semelhança no volume de negócios gerado, cada cliente recorre aos serviços da lavandaria para lavar diferentes tipos de peças, o que pode sugerir diferentes estratégias ao nível da produção ou necessidades específicas em cada caso.

Quanto aos tipos de lavagem, as preferências também variam significativamente. O cliente 34 utiliza principalmente a lavagem ESPECIAL 60° (80%), o cliente 176 prefere a lavagem NORMAL (74%) e o cliente 209 recorre com maior frequência à lavagem NORMAL C/SILICONE (63%). Esta diversidade nas preferências de lavagem indica que a lavandaria deve adaptar suas operações e promover certos tipos de serviços para melhor atender a cada perfil de cliente.

Na análise do total líquido por tipo de lavagem, o algoritmo k-médias sem standardização foi utilizado, e o número ótimo de *clusters* encontrado foi 13. O *cluster* 1, que inclui 13 dos 20 tipos de lavagem, representa 65% dos tipos, mas é responsável por apenas 3,87% do total líquido, o que demonstra que, apesar de uma ampla gama de serviços estar disponível, poucos tipos de lavagem são economicamente relevantes. Os restantes *clusters*, que incluem apenas um tipo de lavagem cada, contribuem significativamente para o total líquido. O tipo de lavagem NORMAL é o mais relevante, com uma participação de 34,05% do total de receitas. Além disso, as lavagens ESPECIAL 60° e NORMAL C/SILICONE também mostraram um crescimento considerável ao longo do tempo, enquanto outros tipos de lavagem, como ANTI-PILLING e C/DISPERSANTE, perderam importância em termos de volume de negócios.

Estes resultados evidenciam a importância de alguns tipos de lavagem específicos para a receita da empresa. Além disso, observou-se que os tipos de lavagem mais

frequentemente utilizados estão associados a tipos de peças responsáveis por maior volume de negócio, como T-SHIRT, POLO, SWEAT e FELPA. Tipos de lavagem como MICRONIZAR e SECO foram utilizados em peças mais diversificadas, o que reforça a importância de oferecer uma gama de lavagens para diferentes tipos de peças e necessidades dos clientes.

Em relação ao total líquido por tipo de peça, o algoritmo k-médias sem standardização identificou dois *clusters*. O *cluster 1* inclui 42 tipos de peças e representa 97,67% do total de peças, enquanto o *cluster 2*, composto apenas pelo tipo T-SHIRT, gera 48,17% do total líquido global. Embora seja apenas um tipo de peça, é responsável por quase metade das receitas da empresa, destacando-se como o principal criador de receita. Este dado demonstra a importância dessa peça para o modelo de negócios da lavanderia, sugerindo que qualquer estratégia de *marketing* ou operação deve dar atenção especial à gestão deste tipo de peça.

Os tipos de lavagem mais utilizados para T-SHIRT são NORMAL C/SILICONE e NORMAL, enquanto outros tipos de lavagem, como ESPECIAL 60°, ANTI-PILLING, e MICRONIZAR, são menos usados. A análise também revelou que os clientes 176, 209, e 34 são os principais responsáveis pelo volume de negócios gerado pela T-SHIRT, corroborando a importância desses clientes identificados na análise por cliente.

A análise de *clustering* aplicada aos dados de faturação da lavanderia revelou informações relevantes sobre o comportamento dos clientes e os serviços que mais contribuem para o volume de negócios. A identificação de clientes mais importantes para as receitas da empresa, como aqueles do *cluster 2*, e a importância de tipos de peças e lavagens específicos, como a T-SHIRT e a lavagem NORMAL, permitem à empresa direcionar melhor suas estratégias comerciais e operacionais.

A previsão de séries temporais aplicada aos dados de faturação da lavanderia industrial, foi realizada com o objetivo de prever o comportamento futuro da empresa, permitindo uma gestão mais informada e orientada por dados. A análise preditiva baseia-se na organização dos dados de faturação em séries temporais, considerando a evolução mensal dos valores ao longo do período de 2018 a 2022. Os dados de 2023 foram utilizados para validar os modelos de previsão desenvolvidos. A análise foi restrita à previsão do total líquido mensal, uma vez que essa variável é fundamental para entender o desempenho financeiro da empresa e prever as receitas futuras.

A metodologia seguida consistiu em duas fases principais: numa primeira fase, os dados foram analisados de forma global, sem segmentações, e numa segunda fase, os dados foram divididos em *clusters* previamente definidos, com base no total líquido por cliente, por tipo de peça e por tipo de lavagem. Essa abordagem permitiu uma análise mais detalhada

dos diferentes fatores que influenciam o volume de faturação da lavandaria, levando em consideração as características específicas de cada agrupamento.

Nove modelos de previsão diferentes foram aplicados, divididos entre métodos simples e métodos regressivos e autorregressivos. Os métodos simples incluíram o Mean, Naive, SNaive e Drift, que são modelos básicos, muitas vezes eficazes para séries temporais simples, mas com limitações para padrões mais complexos. Os métodos regressivos e autorregressivos, por outro lado, são mais sofisticados e permitem uma maior adaptação aos dados, considerando variações sazonais e tendências ao longo do tempo. Foram desenvolvidos quatro modelos regressivos, cada um considerando diferentes combinações de sazonalidade e tendências, tanto lineares como quadráticas, ao longo dos 60 períodos mensais analisados, tendo também sido aplicado o modelo ARIMA.

A aplicação dos modelos de previsão aos dados recolhidos e organizados em séries temporais produziu resultados diferentes em cada caso. De uma forma geral, verifica-se que os modelos de previsão produziram resultados com baixas taxas de erro nos *clusters* com mais importância para a empresa ao nível da faturação. Os *clusters* que representam menores volumes de negócio, tiveram erros maiores ao nível do MAPE.

Os nove modelos de previsão foram aplicados ao total líquido global e às combinações do total líquido por cliente, por tipo de lavagem e por tipo de peça, de acordo com os resultados de agrupamento obtidos através da aplicação dos algoritmos de *clustering*.

Quanto ao total líquido global verifica-se que praticamente todos os modelos produziram previsões com erros abaixo dos 35%, quer no ajuste, quer na previsão. No entanto, os modelos que se destacam com os melhores resultados são o 1 e o ARIMA no ajuste aos dados, com erros de 17% e na previsão os modelos com melhor desempenho são o 3 e 4, com erros de 16%. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são adequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que o modelo Drift produz resultados enviesados.

Os erros pouco significativos permitem utilizar os modelos para a previsão da receita futura da empresa com grande precisão. Isto permite à lavandaria industrial prever a sua receita futura, facilitando a definição de estratégias de investimento e gestão de receita.

Relativamente ao total líquido por cliente, os modelos foram aplicados aos dois *clusters* de clientes. No *cluster* 1, os modelos de previsão tiveram um desempenho heterogéneo, em que os modelos simples produziram erros mais altos e os modelos regressivos e autorregressivos erros menores. O modelo com melhor desempenho no ajuste foi o 1, com um erro de 25%, e na previsão os modelos 3 e 4 com um erro de 26%. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos

regressivos são adequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que nenhum modelo produz resultados enviesados. No *cluster 2* os erros foram menores, com o desempenho dos modelos a melhor em quase todos os casos. Ao nível do ajuste o modelo 1 produziu o menor erro (17%) e ao nível da previsão o Mean apresentou um erro de 23%, com os modelos 3 e 4 a apresentarem erros de 27%. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são adequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que os modelos SNaive e Drift produzem resultados enviesados.

A aplicação de modelos de previsão ao total líquido por cliente, particularmente ao *cluster 2*, constituído pelos 3 principais clientes da lavandaria, e responsável por quase 50% da faturação, permite à empresa, tendo em conta a importância fulcral deste *cluster*, ajustar a sua operação tendo em conta as previsões relativas a estes clientes.

No total líquido por tipo de lavagem, os modelos de previsão foram aplicados aos oito *clusters* obtidos. No *cluster 1*, que representa 13 tipos de lavagem, mas apenas 3,87% do total líquido, os erros foram de uma forma geral muito elevados. No entanto, refere-se que apesar de os MAPE serem elevados os MAE correspondem a poucas dezenas de euros. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são desadequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que o modelo Drift produz resultados enviesados. No *cluster 2* (tipo de lavagem NORMAL, responsável por 34% do total líquido), a aplicação dos modelos de previsão teve um desempenho variável, com alguns dos modelos a produzirem previsões com erros baixos, nomeadamente o modelo 1 no ajuste, com um erro de 20% e o Mean na previsão com um 33% na previsão. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são adequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que os modelos SNaive e Drift produzem resultados enviesados. Os resultados obtidos neste *cluster* permitem que a lavandaria com um erro baixo consiga prever as receitas associadas ao tipo de lavagem mais relevante na operação da lavandaria.

No *cluster 3* do total líquido por tipo de lavagem, os modelos de previsão apresentaram resultados bons, com erros baixos, especialmente ao nível da previsão. No ajuste o modelo ARIMA apresentou o erro mais baixo (32%) e na previsão o modelo 1 apresentou um erro de 20%. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são adequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que os modelos SNaive e Drift produzem resultados enviesados. Quanto ao *cluster 4*, os modelos

obtiveram bons resultados ao nível do ajuste, com o modelo 2 a apresentar um erro de 31%, mas na previsão os erros foram ligeiramente mais elevados, com o método Naive e Drift a apresentarem erro de 45%. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são adequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que nenhum dos modelos produz resultados enviesados. Relativamente ao *cluster* 5, que representa 9% do total líquido, os modelos de previsão apresentam erros elevados, quer no ajuste quer na previsão. Os melhores resultados obtiveram-se com o modelo Drift (64%, mas que se verifica que produz resultados enviesados) no ajuste e com o modelo SNaive (50%), na previsão. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são adequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que o modelo Drift produz resultados enviesados.

As previsões relativas ao *cluster* 6 do total líquido por tipo de lavagem, correspondente à lavagem MICRONIZAR, que representa 7% do total líquido, obtiveram erros elevados quer no ajuste, quer na previsão. Os melhores resultados foram obtidos com o modelo 2, 3 e 4 no ajuste (72%) e com o modelo SNaive na previsão (74%). A análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são adequados, ao modelo ARIMA não foi possível aplicar o teste Box-Ljung e da análise dos resíduos extrai-se que os modelos Drift produzem resultados enviesados. Quanto ao *cluster* 7, tipo de lavagem ANTIPILLING, responsável por 6% do total líquido, os resultados da aplicação dos modelos foram erros muito elevados. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são adequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que os modelos SNaive e Drift produzem resultados enviesados. Por fim, o *cluster* 8, tipo de lavagem C/DISPERSANTE, apresentou resultados da aplicação dos modelos de previsão com erros elevados. O modelo 1 produziu um erro mais baixo quer no ajuste (108%) quer na previsão (62%). No entanto, a análise da significância estatística dos modelos regressivos, indicou que todos os modelos regressivos são desadequados. através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que nenhum dos modelos produz resultados enviesados.

A aplicação dos algoritmos de *clustering* ao total líquido por tipo de peça, resultou em dois *clusters*. O *cluster* 1, que compreende 96,67% dos tipos de peças, apresenta bons resultados ao nível da previsão, com os modelos 1, 2, 3 e 4 a apresentarem erros baixos no ajuste (25%) e na previsão o modelo Mean a apresentar um erro de 21% e o modelo ARIMA com um erro de 25%. A análise da significância estatística dos modelos regressivos, indicou

que todos os modelos regressivos são adequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que o modelo Drift produz resultados enviesados. No caso do *cluster* 2, que inclui apenas o tipo de peça T-SHIRT, responsável por 48,2% do total líquido, os modelos de previsão apresentaram bons resultados, com o modelo 1 a dar um erro de 19% no ajuste e os modelos 3 e 4 a apresentarem erros de 26% na previsão. A análise da significância estatística dos modelos regressivos, indicou que todos os modelos 1 e 2 são adequados e os modelos 3 e 4 são marginalmente desadequados, através do teste de Box-Ljung verificou-se que o modelo ARIMA é adequado e da análise dos resíduos extrai-se que o modelo Drift produz resultados enviesados.

A análise dos resultados da aplicação dos modelos de previsão permite constatar que em grande parte das vezes o modelo que fornece melhores resultados no ajuste, não é o modelo com melhores resultados ao nível da previsão. No entanto, os resultados dos modelos de previsão são em geral muito bons, com erros percentualmente pequenos que permitem com segurança estimar o desempenho futuro da empresa, quer globalmente, quer ao nível dos clientes, lavagens e peças. Os resultados permitiram constatar que os *clusters* com maior representatividade percentual no total líquido apresentam melhores níveis de precisão nos modelos, enquanto os *clusters* com percentagens inferiores a 10% (como os *clusters* 1, 5, 6, 7 e 8) registam os erros mais elevados. Em termos de desempenho, os modelos regressivos e autorregressivos destacam-se no ajuste, enquanto na previsão os resultados são mais heterogêneos, com uma divisão entre estes modelos e os métodos simples.

Em termos globais, pode-se concluir que a aplicação dos modelos de previsão contribui decisivamente para a previsão do desempenho futuro da lavandaria, com os erros a serem baixos nos grupos de clientes, lavagens ou peças que realmente se constituem como os maiores participantes do total líquido da empresa. Estes resultados, permitem apoiar a tomada de decisão, quer ao nível estratégico, quer ao nível operacional e a estimar a receita prevista para o futuro próximo da empresa.

O trabalho de análise de dados realizado, pretendeu contribuir para a transformação da lavandaria industrial em estudo numa empresa orientada por dados. As limitações ao nível de prazo para elaboração desta dissertação não permitem analisar todas as dimensões associadas a esta transformação. No entanto, este foi um passo na direção da implementação desta transformação.

O estudo realizado permitiu identificar *clusters* de clientes, de tipos de lavagens e de tipos de peças, agrupados de acordo com o total líquido gerado por cada um deles. Esta análise permite à empresa conhecer melhor a sua operação, os seus fluxos e tendências de receita e a divisão das mesmas, bem como revelou a existência de grupos cuja similaridade

permite desenvolver estratégias dirigidas em particular a cada um deles. Estas estratégias poderão passar por políticas de *marketing*, por adaptação de preços e pelo desenvolvimento de serviços e ofertas particulares, para cada grupo.

Os grandes contribuintes para a receita da empresa foram identificados, passando do conhecimento empírico, para conhecimento baseado em dados. Esta transformação permite que decisões futuras venham a ser tomadas de forma sustentada, no sentido de maximizarem a capacidade produtiva e em última análise os resultados líquidos da empresa.

Os futuros investimentos da empresa em equipamentos, poderão ter em conta a importância relativa de cada tipo de peça e de cada tipo de lavagem, de forma a adaptar a escolha de cada máquina de lavar, secador, etc. às necessidades efetivas dos principais tipos de lavagem e de peça, realizados na operação da empresa.

Em termos de clientes, a empresa passou a ter um conhecimento sustentado de qual a importância relativa de cada um, devendo fazer uso desta informação no sentido de adaptar as suas políticas de preços e estratégias de *marketing* a cada um dos segmentos, de forma a maximizar a satisfação dos clientes principais e consequentemente melhorar o desempenho global da empresa.

Outra parte deste estudo, prendeu-se com a previsão de receita futura. Esta análise, apoiada em diferentes modelos de previsão de séries temporais, obteve diferentes resultados para cada *cluster* encontrado na análise realizada. Alguns dos resultados obtidos podem ser considerados de elevada precisão, com taxas de erro abaixo dos 20 %. Neste caso, destaca-se a previsão do total líquido global, ou seja, sem qualquer divisão por cliente, tipo de lavagem ou peça, em que os resultados obtidos apresentam erros de 17 % na previsão futura. No entanto, a análise por *clusters* foi mais heterogénea em termos de resultados obtidos, variando desde resultados com baixas taxas de erro, a resultados com taxas de erro muito elevadas. Isto deve-se ao comportamento das séries temporais da variação do total líquido, que, por serem resultados do desempenho real de uma empresa em funcionamento, apresentam grandes variações ao longo do tempo. Algumas destas séries temporais, apresentam comportamentos muito irregulares, o que torna muito difícil o desenvolvimento de modelos de previsão de resultados.

As dificuldades de previsão de algumas das séries temporais, associadas ao comportamento das mesmas, decorrem de diferentes situações, que afetaram e afetam o comportamento da receita da lavandaria. Os efeitos da pandemia provocada pela COVID-19, variações ao nível dos padrões de consumo dos clientes finais, variações ao nível da procura dos clientes, efeitos relacionados com a concorrência do mercado, etc. podem explicar o comportamento errático de algumas séries temporais.

Apesar das dificuldades de previsão em alguns *clusters*, uma grande parte dos resultados permitiu obter previsões com erros baixos. Assim, estes modelos podem ser aplicados à previsão das receitas futuras da empresa. Isto permite apoiar em dados, obtidos de forma científica, decisões sobre investimentos e gestão financeira da empresa.

7.2. Relação com outros estudos

A presente dissertação propôs-se aplicar técnicas de análise de dados com o objetivo de apoiar a transição de uma lavandaria industrial para uma empresa orientada por dados. Os resultados obtidos permitiram validar empiricamente diversos conceitos abordados na revisão de literatura, estabelecendo uma relação entre os fundamentos teóricos e a realidade prática da empresa estudada.

A literatura consultada destaca a crescente relevância do *machine learning* e do *data mining* na indústria têxtil, sobretudo na automatização de processos, controlo de qualidade, e previsão de eventos (Ahmed & Junayed, 2025; Jebbor et al., 2024). Estes conceitos foram materializados neste estudo através da aplicação de algoritmos de *clustering* e de modelos preditivos aos dados de faturação da lavandaria. Os agrupamentos encontrados por cliente, tipo de lavagem e tipo de peça, assim como os padrões de sazonalidade e tendências identificados nas séries temporais, demonstram como estas ferramentas podem produzir conhecimento estruturado a partir de dados operacionais aparentemente dispersos. Assim, confirma-se a aplicabilidade prática das abordagens referidas na literatura ao setor da lavagem e acabamento têxtil.

Autores como Jebbor et al. (2024) salientam o valor da inteligência artificial na tomada de decisão proativa, nomeadamente nas cadeias de abastecimento. Este trabalho reforça essa perspetiva, uma vez que a previsão de receitas futuras, quer a nível global, quer segmentada por *clusters*, permite à empresa planear com maior detalhe os seus investimentos, políticas de marketing e aquisição de equipamentos. A capacidade de antecipar comportamentos de clientes e flutuações na procura reforça a resiliência organizacional e permite uma gestão mais informada, alinhada com os princípios defendidos na literatura.

Tal como abordado por Ersoz (2017) e Makalesi et al. (2024), a análise de dados no setor têxtil pode contribuir para a eficiência operacional, ajudando a identificar fatores críticos, a otimizar recursos e a reduzir riscos. No caso da lavandaria em estudo, a identificação de clientes mais relevantes e de peças e lavagens mais importantes em termos de receita, oferece apoio específico para decisões que promovam a melhoria do desempenho económico e operacional. Ainda que a produtividade individual por trabalhador ou riscos de segurança não tenham sido objeto direto deste trabalho, o potencial para futuras aplicações de *data mining* nestas dimensões foi claramente identificado.

A literatura, nomeadamente Sekhar et al. (2019), destaca a importância de modelos preditivos ajustados ao contexto. Esta conclusão é corroborada pelos resultados do presente estudo, em que os modelos preditivos apresentaram melhor desempenho nos *clusters* mais representativos (com maior peso na faturação) e erros mais elevados nos clusters menos significativos. Isto realça a necessidade de adequar os modelos ao volume e à estabilidade dos dados, reforçando a ideia de que não existe uma solução única universal para a previsão em ambientes complexos e dinâmicos como o setor têxtil.

A correspondência entre os conceitos teóricos analisados na revisão de literatura e os resultados obtidos neste estudo evidencia a relevância prática da ciência de dados aplicada ao setor têxtil. A experiência da Precioso & Morais, Lda. constitui um exemplo real de como técnicas de *clustering* e previsão podem ser integradas na gestão, de forma a promover eficiência, sustentabilidade e vantagem competitiva.

Este trabalho, ainda que limitado, abre portas para futuros desenvolvimentos na empresa, nomeadamente nas áreas de automatização da recolha de dados, monitorização operacional, análise de produtividade e sustentabilidade ambiental, reforçando a importância de uma abordagem integrada e orientada por dados.

7.3. Trabalho futuro

Esta dissertação é por natureza limitada na sua extensão, existindo assim potencial para o alargamento do estudo em trabalhos futuros. Estes, primariamente, poderão passar por completar a análise do pacote de dados recolhido para este trabalho, sendo que a análise foi apenas realizada relativamente ao total líquido. Os dados relativos à quantidade de peças de cada guia de transporte não foram analisados. Assim, a análise desses dados, recorrendo aos métodos apresentados neste estudo, seria de interesse, revelando possíveis novos *clusters* de clientes, lavagens e peças, permitindo obter informações importantes para a gestão operacional da lavandaria, nomeadamente organização de turnos, aquisição de novas máquinas, gestão de prazos de entrega, etc.

Para além da análise complementar aos dados recolhidos, existem outros trabalhos complementares que contribuem para a transformação da empresa em estudo numa empresa orientada por dados. Aspetos como a automatização da recolha de dados, da gestão de encomendas, do armazenamento de dados e posterior análise, que permitam obter informações sobre o desempenho operacional da empresa, tempos de espera, tempos de produção, minimização de desperdícios, etc. De igual forma, a criação de um sistema de apresentação de dados, amigável ao utilizador, que proporcione ao utilizador/decisor a informação trabalhada, atualizada e de forma simples, para apoiar os processos operacionais e de decisão.

A incorporação de técnicas de inteligência artificial representa outra oportunidade para aumentar a eficiência operacional da lavandaria. A aplicação de algoritmos de *machine learning* para análise preditiva pode melhorar a previsão da procura, otimizar a alocação de recursos e antecipar eventuais avarias nos equipamentos, reduzindo os tempos de paragem. Para além disso, a utilização da IA na automatização de processos, como a gestão inteligente de encomendas e o controlo de qualidade, poderá aumentar a precisão e rapidez das operações. A integração de sistemas de IA com interfaces intuitivas permitirá uma tomada de decisão mais informada e célere, reforçando a competitividade da empresa e promovendo uma transformação digital alinhada com as tendências tecnológicas atuais.

Para além do aprofundamento da análise dos dados já recolhidos, existe uma oportunidade significativa para o desenvolvimento de um sistema digital que potencie a transformação da empresa em estudo numa organização orientada por dados. Este sistema teria como principais funcionalidades a automatização da recolha e armazenamento de dados, a gestão integrada de encomendas e lavagens e a disponibilização de análises e indicadores de desempenho operacional em tempo real. Um sistema desta natureza permitiria, por exemplo, analisar de forma simples e imediata os dados relativos à quantidade de peças por guia de transporte, viabilizando novas análises e a identificação de padrões relevantes para a gestão operacional da lavandaria (organização de turnos, planeamento da produção, gestão de prazos, entre outros).

A implementação de um sistema de visualização de dados, com uma interface simples e acessível ao utilizador, permitiria aos decisores aceder a informação tratada e atualizada de forma intuitiva, facilitando os processos de tomada de decisão baseados em dados. Este sistema, embora desenvolvido no contexto de uma lavandaria industrial específica, deverá ser concebido com base numa abordagem genérica e modular, o que abre caminho à sua escalabilidade e comercialização. O sistema poderá ser adaptado com relativa facilidade a outras lavandarias industriais, respondendo às suas especificidades operacionais, mas mantendo a estrutura de recolha, processamento, análise e visualização de dados. Os princípios subjacentes à sua arquitetura deverão ser transversais a diversas áreas de serviços e produção, tais como centros logísticos, empresas de manutenção, ou indústrias com processos repetitivos, de forma a permitir a criação de uma solução digital com valor comercial alargado.

Um sistema como este tem potencial para evoluir de uma ferramenta interna para um produto tecnológico comercializável, integrável em modelos de negócio baseados em software, com serviços associados de implementação, suporte técnico e personalização. O desenvolvimento de funcionalidades adicionais, como a previsão de volumes de trabalho ou a deteção de anomalias com recurso a inteligência artificial e da integração de redes neuronais

artificiais, poderá acrescentar valor ao sistema, reforçando o seu posicionamento como solução avançada de apoio à decisão e à gestão operacional.

Paralelamente, poderá ser também relevante explorar a aplicação de ferramentas de análise de dados e inteligência artificial à gestão de stocks, permitindo uma previsão mais precisa das necessidades de produtos, minimizando ruturas ou excessos de inventário, contribuindo para uma operação mais eficiente e sustentável. Do mesmo modo, a monitorização contínua de indicadores relacionados com o bem-estar dos colaboradores, como níveis de carga de trabalho, pausas, rotatividade e satisfação, poderá ser essencial para promover um ambiente de trabalho mais saudável e equilibrado.

Um outro aspeto a abordar num futuro trabalho, poderá passar pela elaboração de um plano de *marketing* para esta empresa, que tenha em conta os *clusters* obtidos na análise dos dados recolhidos, no sentido de potenciar a relação com os clientes atuais e futuros, maximizando a eficiência e o desempenho da empresa.

Outra direção de desenvolvimento de trabalhos futuros prende-se com a integração de práticas sustentáveis na operação da lavandaria, alinhando a gestão orientada por dados com objetivos ambientais. A análise detalhada do consumo energético, do uso de água e de produtos químicos, bem como a otimização desses recursos através de modelos preditivos e algoritmos de eficiência, poderá contribuir para a redução do impacto ambiental da empresa. Além disso, a implementação de indicadores de sustentabilidade e a monitorização contínua desses parâmetros, apoiadas por sistemas automatizados de recolha e análise de dados, permitirão à empresa adotar medidas proativas para minimizar desperdícios, promover a economia circular e responder às exigências regulatórias e sociais atuais, fortalecendo assim sua responsabilidade ambiental e competitividade no mercado.

Esta dissertação permitiu dar um primeiro passo no sentido de tornar esta empresa numa empresa orientada por dados. A aplicação de técnicas de *data mining* aos dados recolhidos da operação da empresa, permitiu obter um conhecimento mais profundo sobre o desempenho operacional e as características do mesmo e ainda o desenvolvimento de modelos de previsão, passíveis de serem aplicadas no processo de tomada de decisão empresarial, assentando-o em informações obtidas através de um processo de análise de dados.

Apesar de a transformação ser um processo longo, que deve ser desenvolvido em vários vetores e cumprindo várias fases, este estudo serviu de arranque no sentido de permitir que os dados recolhidos diariamente na operação da lavandaria, sejam analisados e contribuam para a melhoria e o desenvolvimento da atividade.

Por mais longa que seja a caminhada, o mais importante é dar o primeiro passo...

8. Bibliografia

- Abdolrasol, M. G. M., Suhail Hussain, S. M., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., Ali, J. A., Mekhilef, S., & Milad, A. (2021). Artificial neural networks based optimization techniques: A review. In *Electronics (Switzerland)* (Vol. 10, Issue 21). MDPI. <https://doi.org/10.3390/electronics10212689>
- Abedinia, A., & Seydi, V. (2024). Building semi-supervised decision trees with semi-cart algorithm. *International Journal of Machine Learning and Cybernetics*, 15(10), 4493–4510. <https://doi.org/10.1007/s13042-024-02161-z>
- Ahmed, T., & Junayed, A. S. M. (2025). Structuring a textile knitting dataset for machine learning and data mining applications. *Data in Brief*, 61, 111873. <https://doi.org/10.1016/j.dib.2025.111873>
- Alves Gomes, M., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and E-Business Management*, 21(3), 527–570. <https://doi.org/10.1007/s10257-023-00640-4>
- American Psychological Association. (2020). *How to Use the Publication Manual 7* (American Psychological Association, Ed.). American Psychological Association.
- Anderson, Carl. (2015). *Creating a data-driven organization: practical advice from the trenches* (T. McGovern, Ed.; First). O'Reilly Media.
- Anderson, D., Sweeney, D., & Williams, T. (2010). Statistics for Business and Economics, Eleventh Edition. *South-Western, Cengage Learning*.
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1), 243–256. <https://doi.org/10.1016/j.patcog.2012.07.021>
- Arora, P., Deepali, & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm for Big Data. *Physics Procedia*, 78, 507–512. <https://doi.org/10.1016/j.procs.2016.02.095>
- Assur, N., & Rowshankish, K. (2022). The data-driven enterprise of 2025. In *McKinsey*.
- Buzrul, S. (2024). On The Pros and Cons of Using Excel for Regression Analysis. *Turkish Journal of Agriculture - Food Science and Technology*, 12(s2), 2234–2241. <https://doi.org/10.24925/turjaf.v12is2.2234-2241.6931>
- Chen, K., Abtahi, F., Carrero, J. J., Fernandez-Llatas, C., & Seoane, F. (2023). Process mining and data mining applications in the domain of chronic diseases: A systematic review. In *Artificial Intelligence in Medicine* (Vol. 144). Elsevier B.V. <https://doi.org/10.1016/j.artmed.2023.102645>

- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 1–24. <https://doi.org/10.7717/PEERJ-CS.623>
- Comissão Europeia. (2020). *COMUNICAÇÃO DA COMISSÃO AO PARLAMENTO EUROPEU, AO CONSELHO, AO COMITÉ ECONÓMICO E SOCIAL EUROPEU E AO COMITÉ DAS REGIÕES - Uma estratégia europeia para os dados*.
- Crummenerl, C., Bohn, U., & Wolf, C. (2023). Boosting change success with data. *Capgemini Invent*.
- Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. In *Renewable and Sustainable Energy Reviews* (Vol. 74, pp. 902–924). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2017.02.085>
- Deldadehasl, M., Karahroodi, H. H., & Haddadian Nekah, P. (2025). Customer Clustering and Marketing Optimization in Hospitality: A Hybrid Data Mining and Decision-Making Approach from an Emerging Economy. *Tourism and Hospitality*, 6(2). <https://doi.org/10.3390/tourhosp6020080>
- Ersoz, F. (2017). *Knowledge Discovery and Data Mining Techniques in Textile Industry*. <https://www.researchgate.net/publication/323279366>
- Esling, P., & Agon, C. (2012). Time-series data mining. *ACM Computing Surveys*, 45(1). <https://doi.org/10.1145/2379776.2379788>
- Ezugwu, A. E., Shukla, A. K., Agbaje, M. B., Oyelade, O. N., José-García, A., & Agushaka, J. O. (2021). Automatic clustering algorithms: a systematic review and bibliometric analysis of relevant literature. In *Neural Computing and Applications* (Vol. 33, Issue 11, pp. 6247–6306). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s00521-020-05395-4>
- Garibaldi, U., Costantini, D., Donadio, S., & Viarengo, P. (2006). Herding and clustering in economics: The Yule-Zipf-Simon model. *Computational Economics*, 27(1), 115–134. <https://doi.org/10.1007/s10614-005-9018-y>
- Ghazal, T. M., Hussain, M. Z., Said, R. A., Nadeem, A., Hasan, M. K., Ahmad, M., Khan, M. A., & Naseem, M. T. (2021). Performances of k-means clustering algorithm with different distance metrics. *Intelligent Automation and Soft Computing*, 30(2), 735–742. <https://doi.org/10.32604/iasc.2021.019067>
- Gladju, J., Kamalam, B. S., & Kanagaraj, A. (2022). Applications of data mining and machine learning framework in aquaculture and fisheries: A review. In *Smart Agricultural Technology* (Vol. 2). Elsevier B.V. <https://doi.org/10.1016/j.atech.2022.100061>

- Herman, E., Zsido, K. E., & Fenyves, V. (2022). Cluster Analysis with K-Mean versus K-Medoid in Financial Performance Evaluation. *Applied Sciences (Switzerland)*, 12(16). <https://doi.org/10.3390/app12167985>
- Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. In *Briefings in Bioinformatics* (Vol. 24, Issue 2). Oxford University Press. <https://doi.org/10.1093/bib/bbad002>
- Huang, J. J., Tzeng, G. H., & Ong, C. S. (2007). Marketing segmentation using support vector clustering. *Expert Systems with Applications*, 32(2), 313–317. <https://doi.org/10.1016/j.eswa.2005.11.028>
- Hupperz, M., Möller, F., & Otto, B. (2021). What is a Data-Driven Organization? *Twenty-Seventh Americas Conference on Information Systems*. <https://www.researchgate.net/publication/351282206>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (Second). OTexts.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Javaid, H. A. (2024). *Improving Fraud Detection and Risk Assessment in Financial Service using Predictive Analytics and Data Mining*. <https://ijstpublication.com>
- Jebbor, I., Benmamoun, Z., & Hachimi, H. (2024). Forecasting supply chain disruptions in the textile industry using machine learning: A case study. *Ain Shams Engineering Journal*, 15(12). <https://doi.org/10.1016/j.asej.2024.103116>
- Kamal Kaur, N., Kaur, U., & Singh, D. (2014). K-Medoid Clustering Algorithm-A Review. *International Journal of Computer Application and Technology (IJCAT)*, 1(1). www.technopublications.com
- Kamel, A. R., & Abonazel, M. R. (2023). A Simple Introduction to Regression Modeling using R. *Computational Journal of Mathematical and Statistical Sciences*, 2(1), 52–79. <https://doi.org/10.21608/cjmss.2023.189834.1002>
- Khan, M. A., Saqib, S., Alyas, T., Ur Rehman, A., Saeed, Y., Zeb, A., Zareei, M., & Mohamed, E. M. (2020a). Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning. *IEEE Access*, 8, 116013–116023. <https://doi.org/10.1109/ACCESS.2020.3003790>
- Khan, M. A., Saqib, S., Alyas, T., Ur Rehman, A., Saeed, Y., Zeb, A., Zareei, M., & Mohamed, E. M. (2020b). Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning. *IEEE Access*, 8, 116013–116023. <https://doi.org/10.1109/ACCESS.2020.3003790>

- Kolkova, A. (2020). The application of forecasting sales of services to increase business competitiveness. *Journal of Competitiveness*, 12(2), 90–105. <https://doi.org/10.7441/joc.2020.02.06>
- Kumar, S. (2020). Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm. *Journal of Artificial Intelligence and Capsule Networks*, 2(1), 53–63. <https://doi.org/10.36548/jaicn.2020.3.006>
- Lee, S., Kim, J., Hwang, J., Lee, E. J., Lee, K. J., Oh, J., Park, J., & Heo, T. Y. (2020). Clustering of time series water quality data using dynamic time warping: A case study from the Bukhan River water quality monitoring network. *Water (Switzerland)*, 12(9). <https://doi.org/10.3390/w12092411>
- Li, Q., Zhang, X., Ma, T., Liu, D., Wang, H., & Hu, W. (2022). A Multi-step ahead photovoltaic power forecasting model based on TimeGAN, Soft DTW-based K-medoids clustering, and a CNN-GRU hybrid neural network. *Energy Reports*, 8, 10346–10362. <https://doi.org/10.1016/j.egyr.2022.08.180>
- Mahdiraji, H. A., Kazimieras Zavadskas, E., Kazeminia, A., & Abbasi Kamardi, A. A. (2019). Marketing strategies evaluation based on big data analysis: a CLUSTERING-MCDM approach. *Economic Research-Ekonomska Istraživanja*, 32(1), 2882–2898. <https://doi.org/10.1080/1331677X.2019.1658534>
- Makalesi, A., Tunçman, B., Gündüz, T., & Eroğlu, D. Y. (2024). Evaluation of Risk Factors Causing Occupational Accidents in the Textile Sector Using Data Mining Methods. *European Journal of Science and Technology*, 53, 84–96. <https://doi.org/10.5281/zenodo.10646656>
- Makridakis, S., Hogarth, R. M., & Gaba, A. (2009). Forecasting and uncertainty in the economic and business world. *International Journal of Forecasting*, 25(4), 794–812. <https://doi.org/10.1016/j.ijforecast.2009.05.012>
- McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research*, 60(6), 656–662. <https://doi.org/10.1016/j.jbusres.2006.06.015>
- Milanović, M., & Stamenković, M. (2016). CHAID Decision Tree: Methodological Frame and Application. *Economic Themes*, 54(4), 563–586. <https://doi.org/10.1515/ethemes-2016-0029>
- Mondal, P., Shit, L., & Goswami, S. (2014). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications*, 4(2), 13–29. <https://doi.org/10.5121/ijcsea.2014.4202>
- Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022a). Fundamentals of Artificial Neural Networks and Deep Learning. In *Multivariate Statistical Machine Learning*

- Methods for Genomic Prediction* (pp. 379–425). Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_10
- Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022b). Support Vector Machines and Support Vector Regression. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (pp. 337–378). Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_9
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Time Series Analysis and Forecasting* (D. Balding, N. Cressie, G. Fitzmaurice, & G. Gives, Eds.; Second). John Wiley & Sons, Inc.
- Muller, M. (2007). Dynamic Time Warping. In *Information Retrieval for Music and Motion* (pp. 69–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74048-3_4
- Regin, R., & Suman Rajest, S. (2023). A Review of Secure Neural Networks and Big Data Mining Applications in Financial Risk Assessment CENTRAL ASIAN JOURNAL OF INNOVATIONS ON TOURISM MANAGEMENT AND FINANCE. CENTRAL ASIAN JOURNAL OF INNOVATIONS ON TOURISM MANAGEMENT AND FINANCE. <http://www.centralasianstudies.orghttps://cajitmf.centralasianstudies.org>
- Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. da F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PLoS ONE*, 14(1). <https://doi.org/10.1371/journal.pone.0210236>
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. In *Journal of Computational and Applied Mathematics* (Vol. 20).
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. <https://doi.org/10.58496/bjml/2024/007>
- Salvador, S., & Chan, P. (2007). Toward accurate dynamic time warping in linear time and space. In *Intelligent Data Analysis* (Vol. 11). IOS Press.
- Saputra, D. M., Saputra, D., & Oswari, L. D. (2020). Advances in Intelligent Systems Research. *SICONIAN 2019*, 172.
- Sekhar, S., Lakshmi, D. C., Das, S., Ketan, B., & Editors, P. (2019). *Advances in Intelligent Systems and Computing 1056*. <http://www.springer.com/series/11156>
- Sheth, J. (2021). New areas of research in marketing strategy, consumer behavior, and marketing analytics: the future is bright. *Journal of Marketing Theory and Practice*, 29(1), 3–12. <https://doi.org/10.1080/10696679.2020.1860679>
- Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6). <https://doi.org/10.3390/e23060759>

- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Statista. (2024). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025*. <https://www.statista.com/statistics/871513/worldwide-data-created/>
- Sweeney, C., Bessa, R. J., Browell, J., & Pinson, P. (2020). The future of forecasting for renewable energy. In *Wiley Interdisciplinary Reviews: Energy and Environment* (Vol. 9, Issue 2). John Wiley and Sons Ltd. <https://doi.org/10.1002/wene.365>
- Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). *Introduction to data mining* (C. Pellegrino, Ed.; Second). Pearson Education Limited.
- Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
- Tsiu, S. V., Ngoben, M., Mathabela, L., & Thango, B. (2025). Applications and Competitive Advantages of Data Mining and Business Intelligence in SMEs Performance: A Systematic Review. *Businesses*, 5(2), 22. <https://doi.org/10.3390/businesses5020022>
- Valkenborg, D., Rousseau, A. J., Geubbelmans, M., & Burzykowski, T. (2023). Support vector machines. In *American Journal of Orthodontics and Dentofacial Orthopedics* (Vol. 164, Issue 5, pp. 754–757). Elsevier Inc. <https://doi.org/10.1016/j.ajodo.2023.08.003>
- Wegmann, M., Dominique, Z., Hillenbrand, J., & Fleischer, J. (2021). *Wegmann et al.: A review of systematic selection of clustering algorithms and their evaluation*.
- Wegmann, M., Zipperling, D., Hillenbrand, J., & Fleischer, J. (2021). *A review of systematic selection of clustering algorithms and their evaluation*.
- Xu, D., & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, 2(2), 165–193. <https://doi.org/10.1007/s40745-015-0040-1>
- Yan, Y. (2022). *Influencing Factors of Housing Price in New York-analysis: Based on Excel Multi-regression Model*. 1005–1009. <https://doi.org/10.5220/0011362000003440>
- Zhang, J., Johnstone, M., Le, V., Khan, B., Hosen, A., Creighton, D., Carney, J., Wilson, A., & Lynch, M. (2021). Dynamic time warp-based clustering: Application of machine learning algorithms to simulation input modelling. *Expert Systems with Applications*, 186. <https://doi.org/10.1016/j.eswa.2021.115684>

9. Apêndices

9.1. Resultados *Clustering*

9.1.1. Total líquido por Cliente

Clientes							
Distância Euclidiana				Dynamic Time Warping			
Estandarizado		Não estandarizado		Estandarizado		Não estandarizado	
nº ot clus	Coef Silh	nº ot clus	Coef Silh	nº ot clus	Coef Silh	nº ot clus	Coef Silh
2	0,9429	2	0,9307	2	0,8904	2	0,8777
Clientes	Cluster	Clientes	Cluster	Clientes	Cluster	Clientes	Cluster
34	2	34	2	34	2	34	2
176	2	176	2	176	2	176	2
209	2	209	2	209	2	209	2
63	2	63	1	63	1	63	1
55	1	55	1	55	2	55	1
196	1	196	1	196	1	196	1
64	1	64	1	64	1	64	1
119	1	119	1	119	1	119	1
254	1	254	1	254	1	254	1
123	1	123	1	123	1	123	1
112	1	112	1	112	1	112	1
62	1	62	1	62	1	62	1
175	1	175	1	175	1	175	1
174	1	174	1	174	1	174	1
247	1	247	1	247	1	247	1
32	1	32	1	32	1	32	1
169	1	169	1	169	1	169	1
129	1	129	1	129	1	129	1
141	1	141	1	141	1	141	1
232	1	232	1	232	1	232	1
167	1	167	1	167	1	167	1
193	1	193	1	193	1	193	1
178	1	178	1	178	1	178	1
155	1	155	1	155	1	155	1
250	1	250	1	250	1	250	1
185	1	185	1	185	1	185	1
259	1	259	1	259	1	259	1
264	1	264	1	264	1	264	1
218	1	218	1	218	1	218	1
127	1	127	1	127	1	127	1
159	1	159	1	159	1	159	1
204	1	204	1	204	1	204	1
163	1	163	1	163	1	163	1
213	1	213	1	213	1	213	1

265	1	265	1	265	1	265	1
110	1	110	1	110	1	110	1
149	1	149	1	149	1	149	1
171	1	171	1	171	1	171	1
241	1	241	1	241	1	241	1
211	1	211	1	211	1	211	1
230	1	230	1	230	1	230	1
236	1	236	1	236	1	236	1
253	1	253	1	253	1	253	1
194	1	194	1	194	1	194	1
261	1	261	1	261	1	261	1
184	1	184	1	184	1	184	1
248	1	248	1	248	1	248	1
94	1	94	1	94	1	94	1
205	1	205	1	205	1	205	1
266	1	266	1	266	1	266	1
206	1	206	1	206	1	206	1
92	1	92	1	92	1	92	1
255	1	255	1	255	1	255	1
225	1	225	1	225	1	225	1
251	1	251	1	251	1	251	1
208	1	208	1	208	1	208	1
217	1	217	1	217	1	217	1
256	1	256	1	256	1	256	1
182	1	182	1	182	1	182	1
257	1	257	1	257	1	257	1
214	1	214	1	214	1	214	1
222	1	222	1	222	1	222	1
271	1	271	1	271	1	271	1
249	1	249	1	249	1	249	1
135	1	135	1	135	1	135	1
244	1	244	1	244	1	244	1
245	1	245	1	245	1	245	1
240	1	240	1	240	1	240	1
273	1	273	1	273	1	273	1
242	1	242	1	242	1	242	1
215	1	215	1	215	1	215	1
258	1	258	1	258	1	258	1
246	1	246	1	246	1	246	1
269	1	269	1	269	1	269	1
252	1	252	1	252	1	252	1
75	1	75	1	75	1	75	1
190	1	190	1	190	1	190	1
237	1	237	1	237	1	237	1

277	1	277	1	277	1	277	1
263	1	263	1	263	1	263	1
22	1	22	1	22	1	22	1
9	1	9	1	9	1	9	1
233	1	233	1	233	1	233	1
272	1	272	1	272	1	272	1
59	1	59	1	59	1	59	1
243	1	243	1	243	1	243	1
279	1	279	1	279	1	279	1
229	1	229	1	229	1	229	1
262	1	262	1	262	1	262	1
164	1	164	1	164	1	164	1
172	1	172	1	172	1	172	1
274	1	274	1	274	1	274	1
260	1	260	1	260	1	260	1
281	1	281	1	281	1	281	1
282	1	282	1	282	1	282	1

9.1.2. Total líquido por Tipo de Lavagem

Tipo de Lavagem							
Distância Euclidiana				Dynamic Time Warping			
Estandarizado		Não estandardizado		Estandarizado		Não estandardizado	
nº ot clus	Coef Silh	nº ot clus	Coef Silh	nº ot clus	Coef Silh	nº ot clus	Coef Silh
8	0,949	8	0,9553	8	0,9089	8	0,8998
Lavagem	Cluste r	Lavagem	Cluste r	Lavagem	Cluste r	Lavagem	Cluste r
NORMAL	2	NORMAL	2	NORMAL	2	NORMAL	2
ESPECIAL 60°	3	ESPECIAL 60°	3	ESPECIAL 60°	4	ESPECIAL 60°	4
NORMAL C/SILICONE	4	NORMAL C/SILICONE	4	NORMAL C/SILICONE	7	NORMAL C/SILICONE	7
SECO	5	SECO	5	SECO	3	SECO	3
MICRONIZAR	6	MICRONIZAR	6	MICRONIZAR	6	MICRONIZAR	5
ANTI-PILLING	7	ANTI-PILLING	7	ANTI-PILLING	5	ANTI-PILLING	8
C/DISPERSANTE	8	C/DISPERSANTE	8	C/DISPERSANTE	8	C/DISPERSANTE	6
STONE-WASH	1	STONE-WASH	1	STONE-WASH	1	STONE-WASH	1
TUMBLER	1	TUMBLER	1	TUMBLER	1	TUMBLER	1
ENZIMÁTICA C/SILICONE	1	ENZIMÁTICA C/SILICONE	1	ENZIMÁTICA C/SILICONE	1	ENZIMÁTICA C/SILICONE	1
ANTI-OLEO	1	ANTI-OLEO	1	ANTI-OLEO	1	ANTI-OLEO	1
ENZIMÁTICA	1	ENZIMÁTICA	1	ENZIMÁTICA	1	ENZIMÁTICA	1
60°	1	60°	1	60°	1	60°	1
ANTI-PILLING C/SILICONE	1	ANTI-PILLING C/SILICONE	1	ANTI-PILLING C/SILICONE	1	ANTI-PILLING C/SILICONE	1
SAND WASH	1	SAND WASH	1	SAND WASH	1	SAND WASH	1
NORMAL AMOSTRAS	1	NORMAL AMOSTRAS	1	NORMAL AMOSTRAS	1	NORMAL AMOSTRAS	1
PROGRAMA 40	1	PROGRAMA 40	1	PROGRAMA 40	1	PROGRAMA 40	1
ANTI-ESTÁTICO	1	ANTI-ESTÁTICO	1	ANTI-ESTÁTICO	1	ANTI-ESTÁTICO	1
ANTI-CORROSÃO	1	ANTI-CORROSÃO	1	ANTI-CORROSÃO	1	ANTI-CORROSÃO	1
FORTE	1	FORTE	1	FORTE	1	FORTE	1

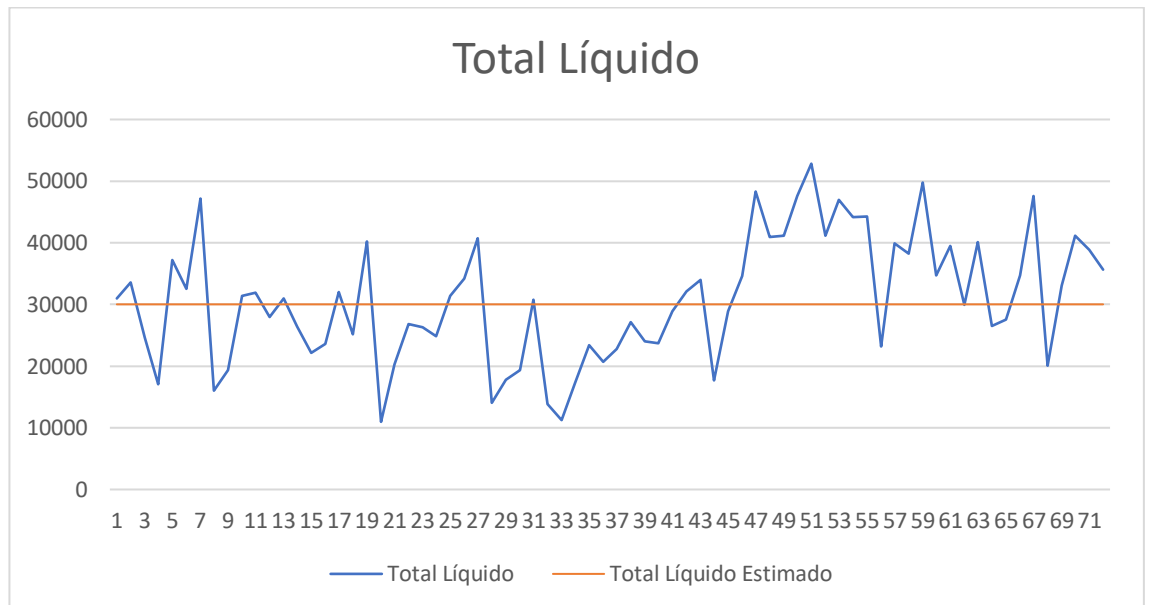
9.1.3. Total Líquido por Tipo de Peça

Tipo de Peça							
Distância Euclidiana				Dynamic Time Warping			
Estandarizado		Não estandarizado		Estandarizado		Não estandarizado	
nº ot clus	Coef Silh	nº ot clus	Coef Silh	nº ot clus	Coef Silh	nº ot clus	Coef Silh
2	0,9814	2	0,9869	2	0,9504	2	0,9509
Peça	Cluster	Peça	Cluster	Peça	Cluster	Peça	Cluster
T-SHIRT	2	T-SHIRT	2	T-SHIRT	2	T-SHIRT	2
POLO	1	POLO	1	POLO	1	POLO	1
SWEAT	1	SWEAT	1	SWEAT	1	SWEAT	1
FELPA	1	FELPA	1	FELPA	1	FELPA	1
NAO DEFINIDO	1	NAO DEFINIDO	1	NAO DEFINIDO	1	NAO DEFINIDO	1
CALÇA	1	CALÇA	1	CALÇA	1	CALÇA	1
CAMISOLA	1	CAMISOLA	1	CAMISOLA	1	CAMISOLA	1
VESTIDO	1	VESTIDO	1	VESTIDO	1	VESTIDO	1
CASACO	1	CASACO	1	CASACO	1	CASACO	1
CAMISA	1	CAMISA	1	CAMISA	1	CAMISA	1
CREPOM	1	CREPOM	1	CREPOM	1	CREPOM	1
SINGLET	1	SINGLET	1	SINGLET	1	SINGLET	1
PAINEIS	1	PAINEIS	1	PAINEIS	1	PAINEIS	1
GOLAS	1	GOLAS	1	GOLAS	1	GOLAS	1
CALÇÃO	1	CALÇÃO	1	CALÇÃO	1	CALÇÃO	1
SAIA	1	SAIA	1	SAIA	1	SAIA	1
TIRAS	1	TIRAS	1	TIRAS	1	TIRAS	1
TOP	1	TOP	1	TOP	1	TOP	1
AMOSTRAS	1	AMOSTRAS	1	AMOSTRAS	1	AMOSTRAS	1
MACACÃO	1	MACACÃO	1	MACACÃO	1	MACACÃO	1
TUNICA	1	TUNICA	1	TUNICA	1	TUNICA	1
PANOS	1	PANOS	1	PANOS	1	PANOS	1
BODY	1	BODY	1	BODY	1	BODY	1
MÁSCARAS	1	MÁSCARAS	1	MÁSCARAS	1	MÁSCARAS	1
HOODIES	1	HOODIES	1	HOODIES	1	HOODIES	1
SUJOS	1	SUJOS	1	SUJOS	1	SUJOS	1
MEIAS	1	MEIAS	1	MEIAS	1	MEIAS	1
CORTINAS	1	CORTINAS	1	CORTINAS	1	CORTINAS	1
BLAZER	1	BLAZER	1	BLAZER	1	BLAZER	1
PIJAMAS	1	PIJAMAS	1	PIJAMAS	1	PIJAMAS	1
BLUSA	1	BLUSA	1	BLUSA	1	BLUSA	1
LEGGINS	1	LEGGINS	1	LEGGINS	1	LEGGINS	1
TUBO	1	TUBO	1	TUBO	1	TUBO	1
ROBE	1	ROBE	1	ROBE	1	ROBE	1
MALHAS	1	MALHAS	1	MALHAS	1	MALHAS	1
FITAS	1	FITAS	1	FITAS	1	FITAS	1
CUECA	1	CUECA	1	CUECA	1	CUECA	1
PUFS	1	PUFS	1	PUFS	1	PUFS	1
RENDAS	1	RENDAS	1	RENDAS	1	RENDAS	1
CORDÃO	1	CORDÃO	1	CORDÃO	1	CORDÃO	1
CAPA	1	CAPA	1	CAPA	1	CAPA	1
LENÇOS	1	LENÇOS	1	LENÇOS	1	LENÇOS	1
COLETES	1	COLETES	1	COLETES	1	COLETES	1

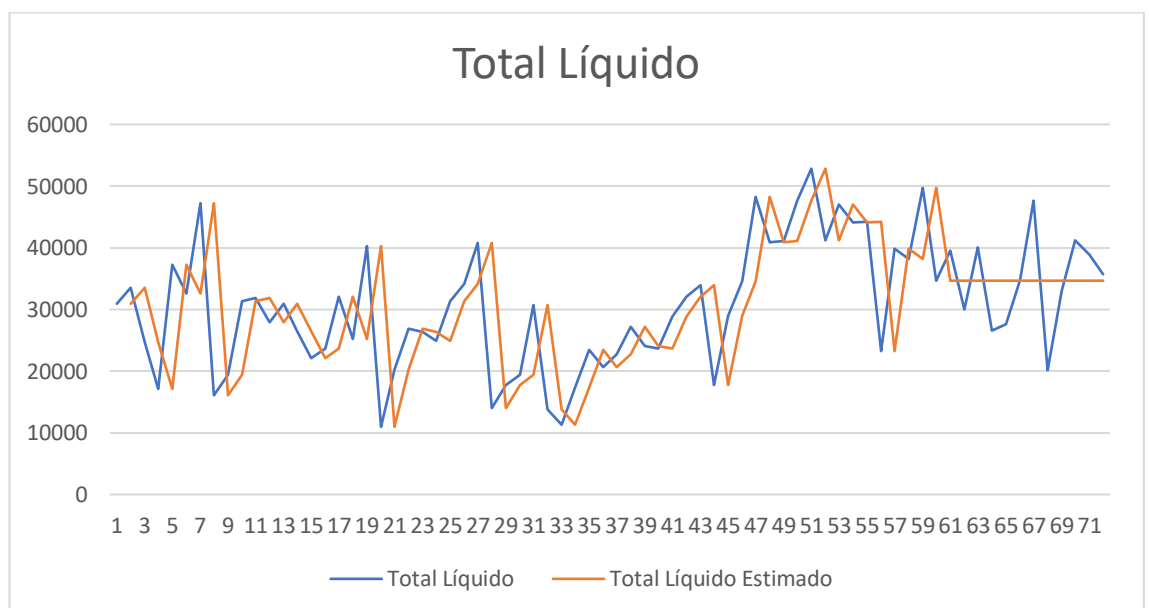
9.2. Gráficos das previsões de séries temporais

9.2.1. Total Líquido Global

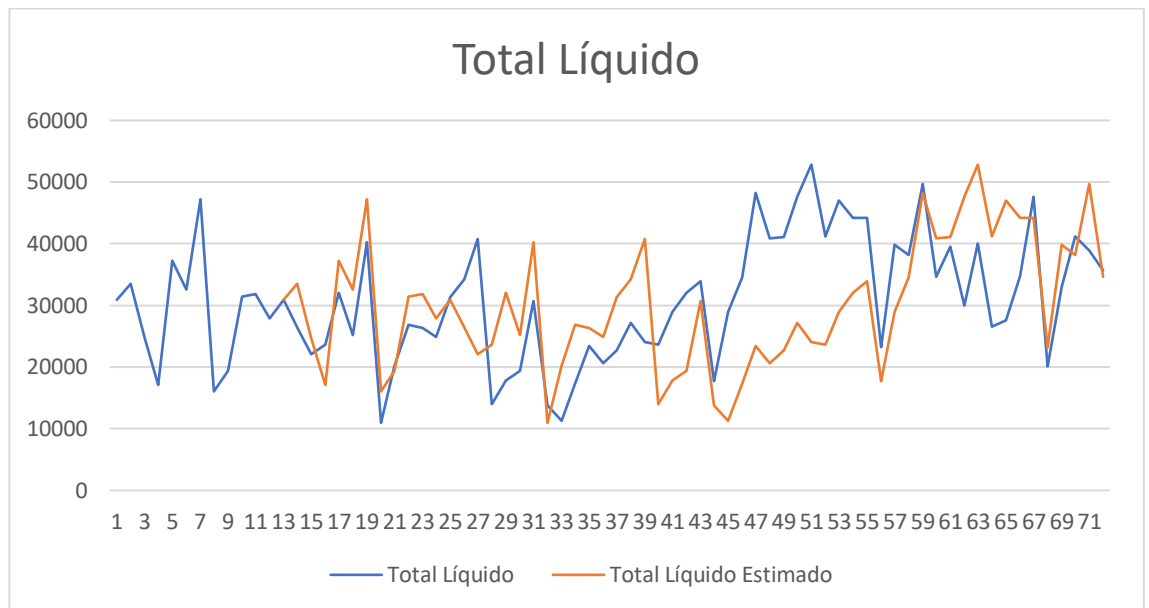
9.2.1.1. Métodos simples - Mean



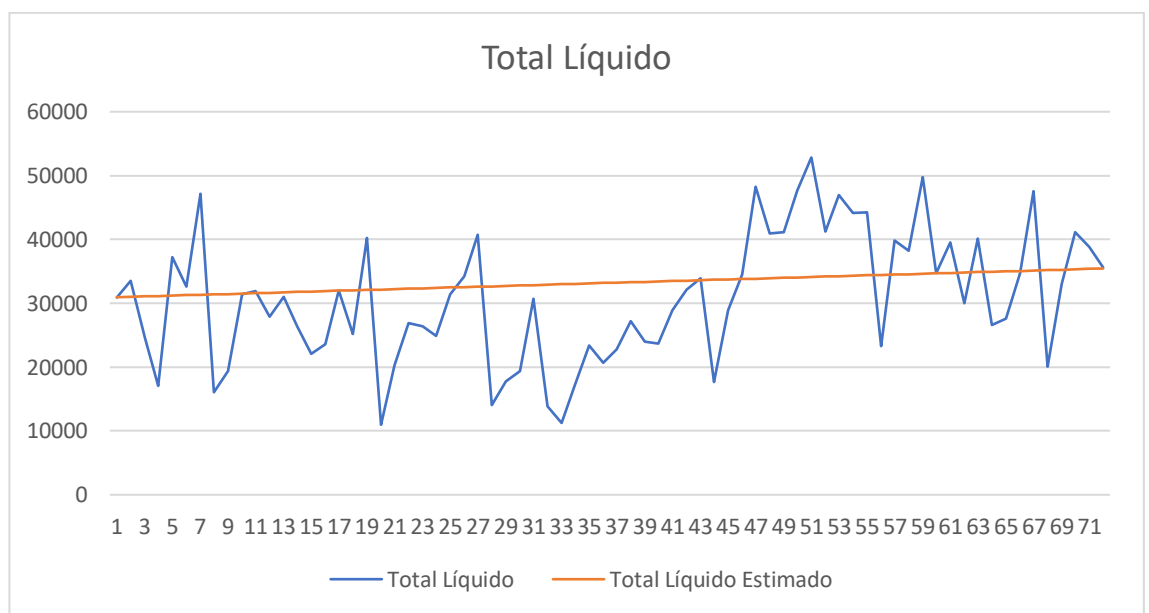
9.2.1.2. Métodos simples – Naive



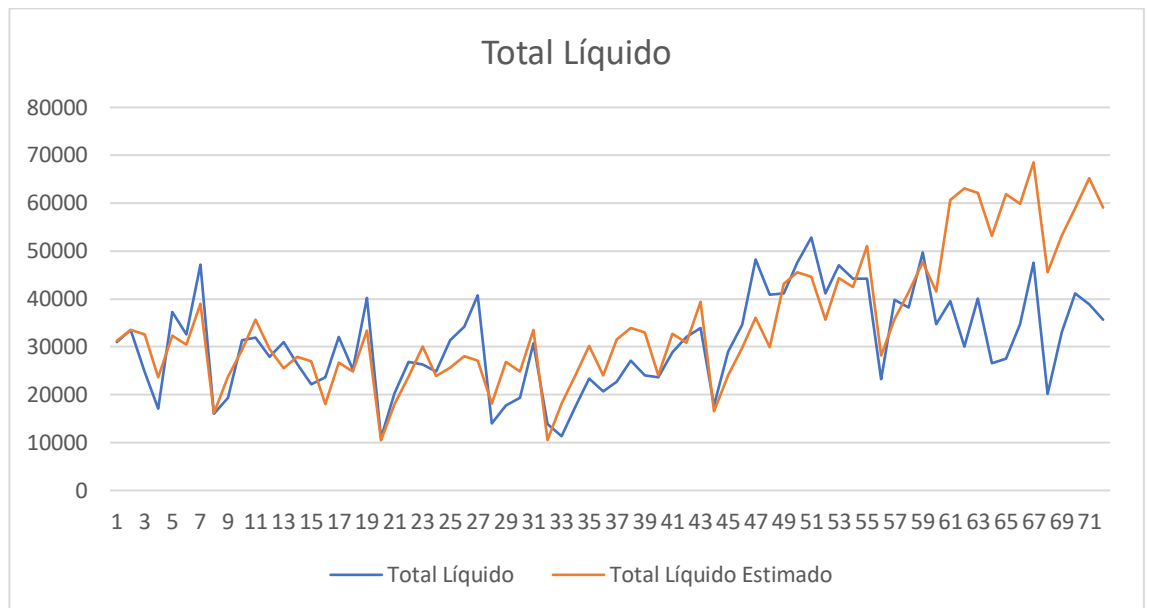
9.2.1.3. Métodos simples – SNaive



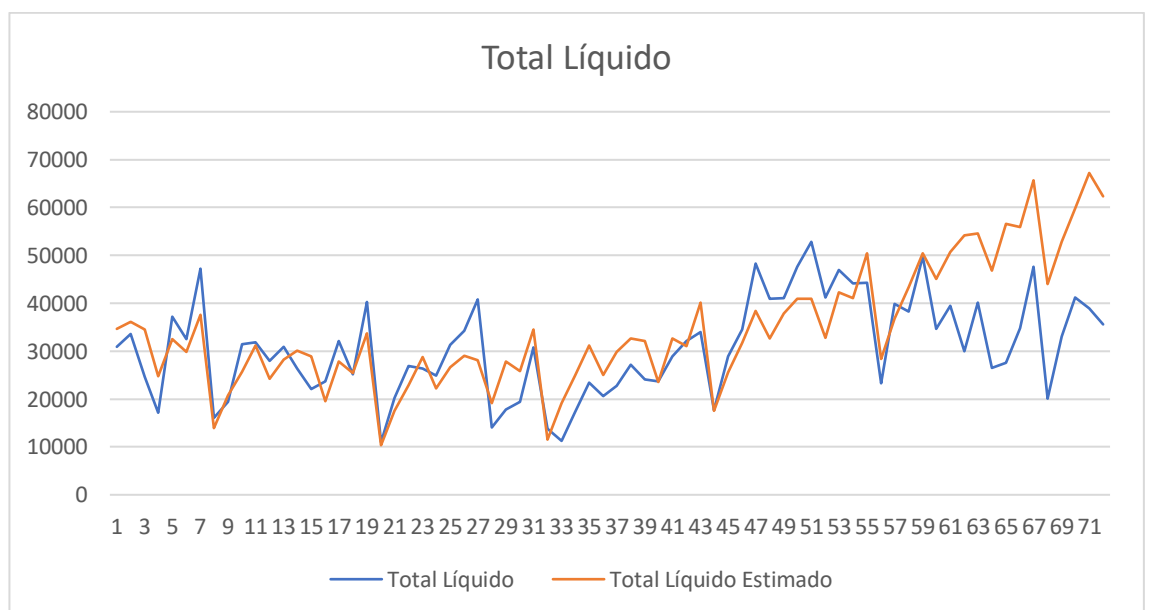
9.2.1.4. Métodos simples – Drift



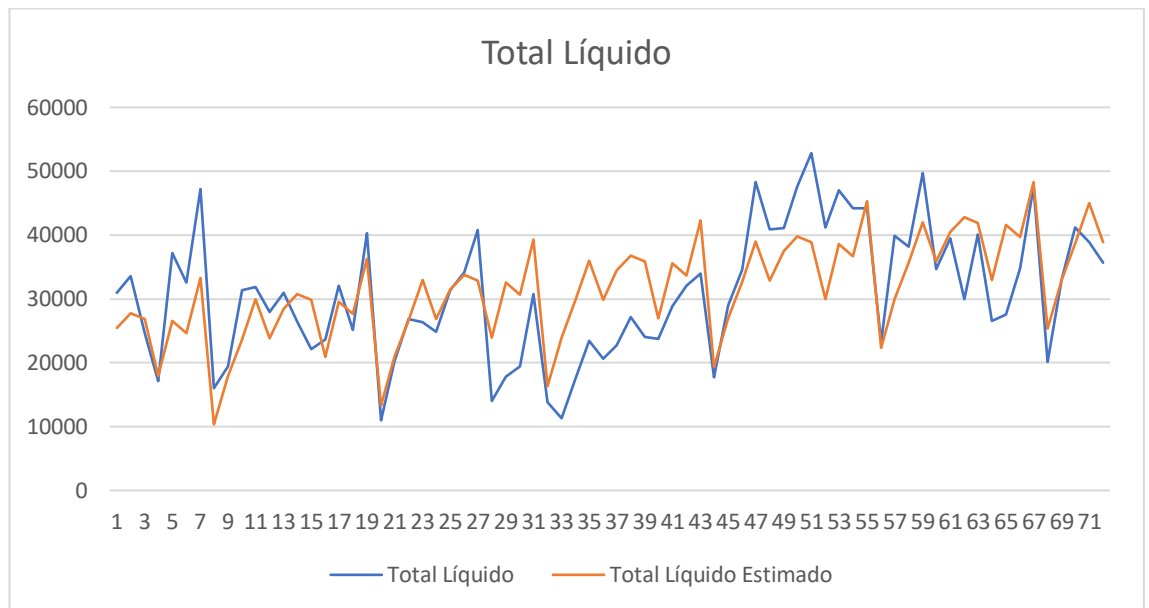
9.2.1.5. Métodos regressivos - Modelo 1



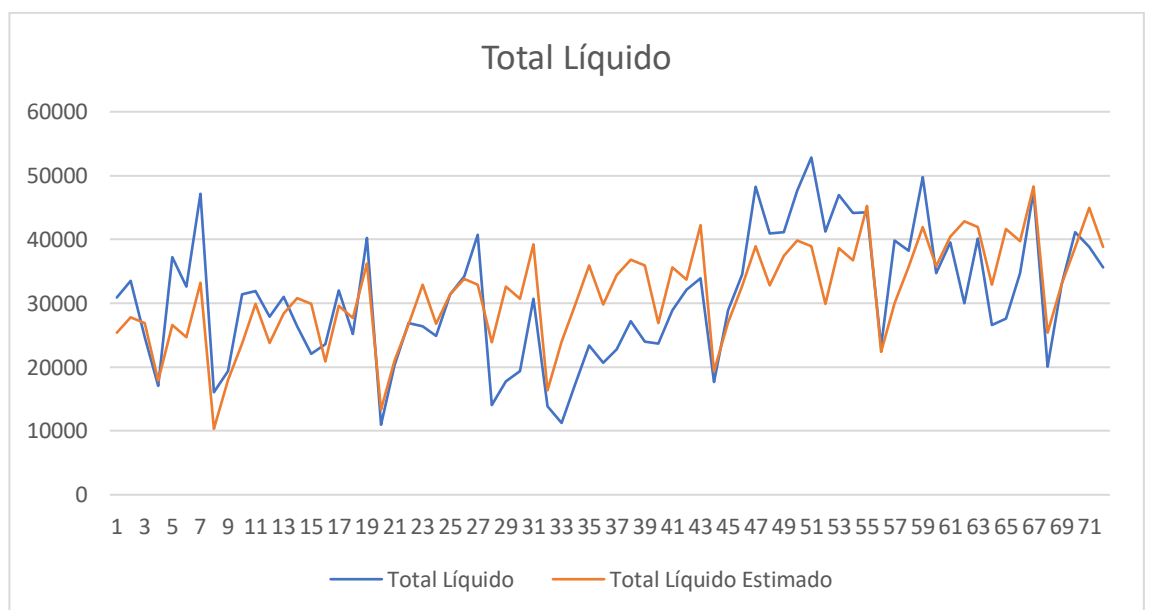
9.2.1.6. Métodos regressivos - Modelo 2



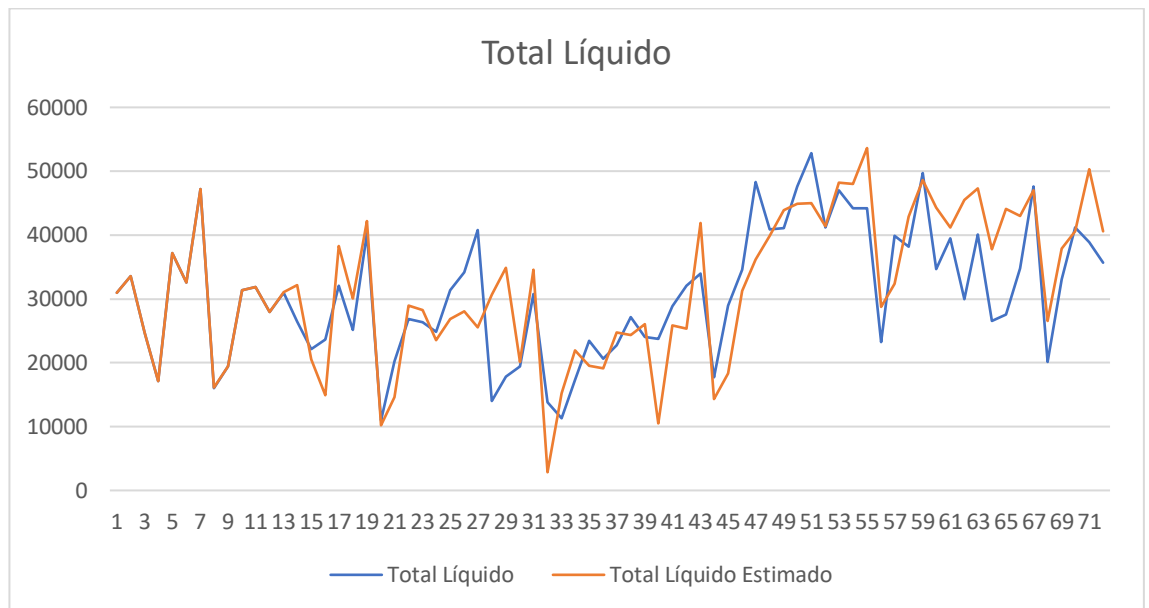
9.2.1.7. Métodos regressivos - Modelo 3



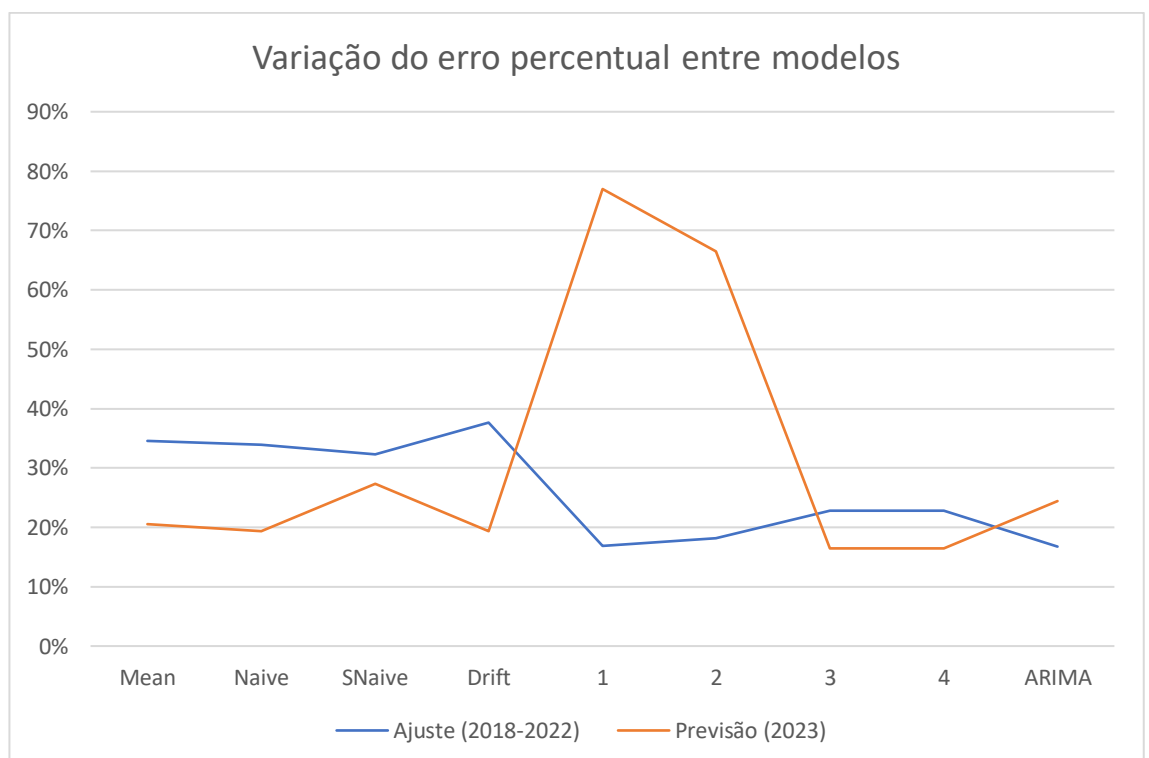
9.2.1.8. Métodos regressivos - Modelo 4



9.2.1.9. ARIMA

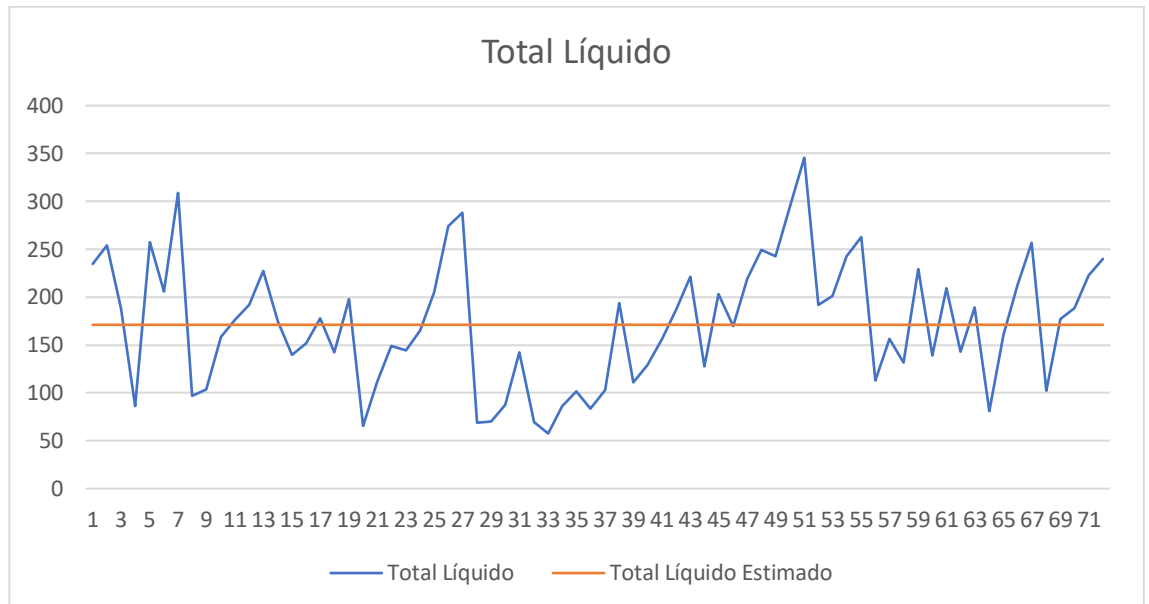


9.2.1.10. Representação gráfica dos erros dos modelos

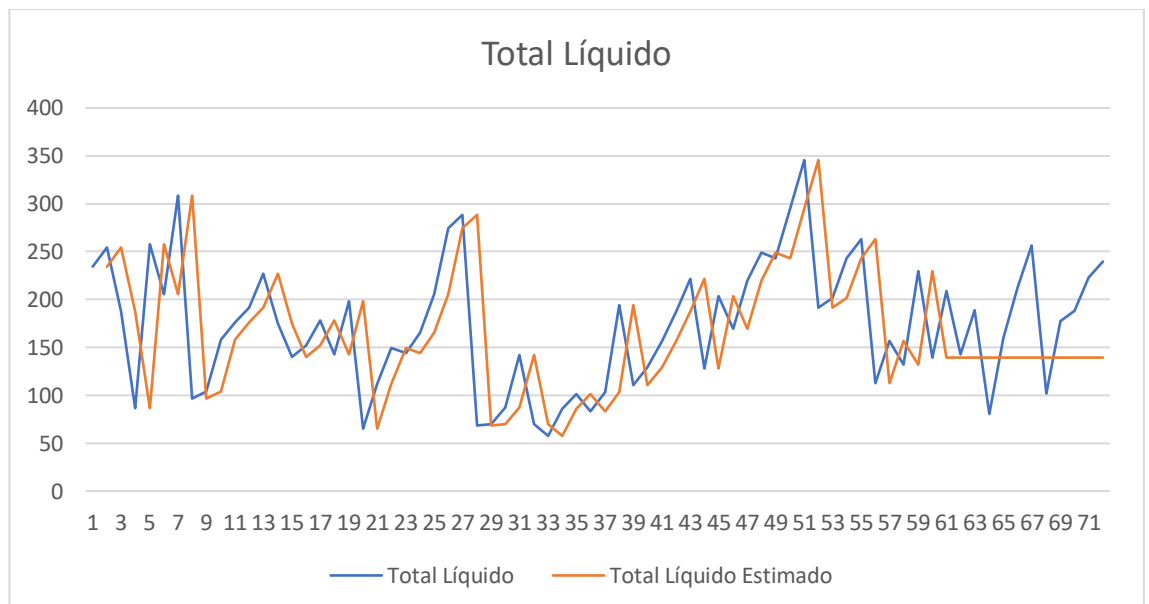


9.2.2. Total Líquido por Cliente – *Cluster 1*

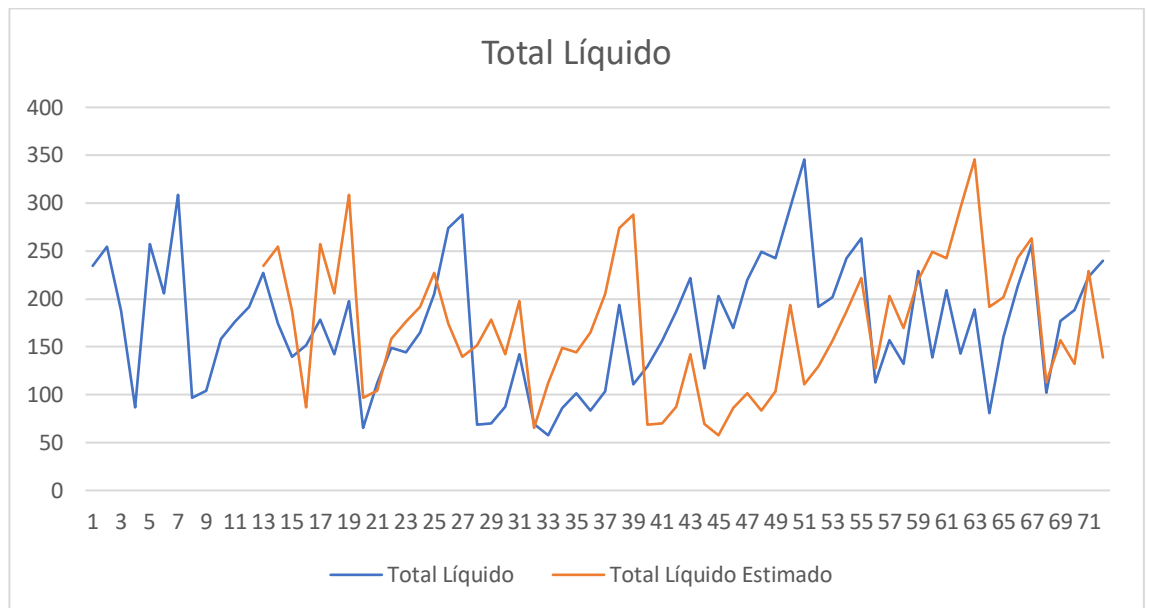
9.2.2.1. Métodos simples - Mean



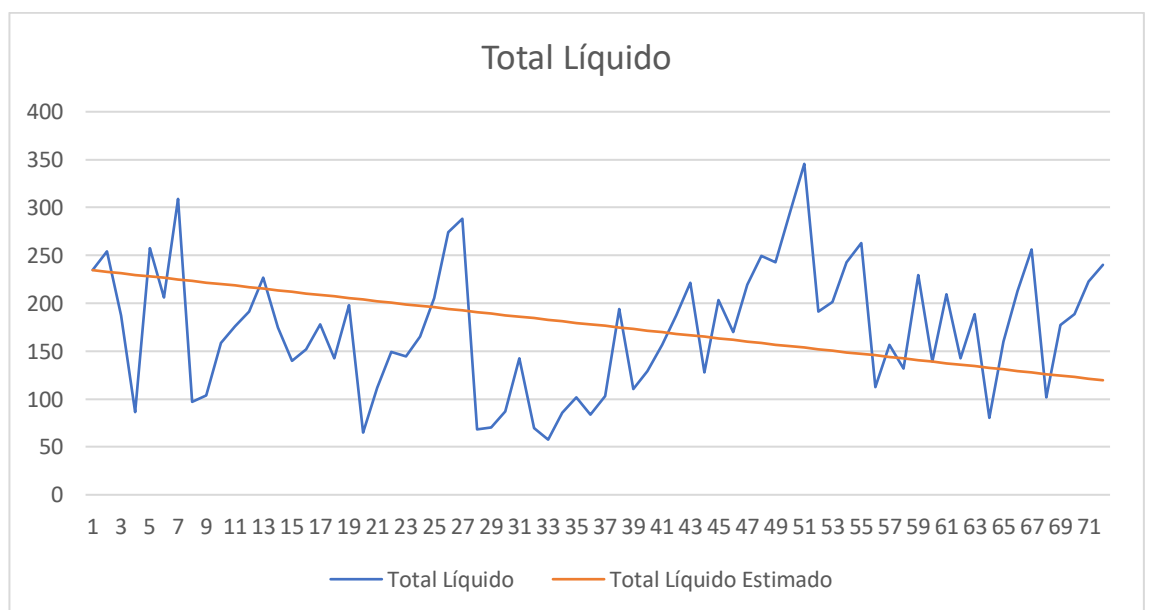
9.2.2.2. Métodos simples - Naive



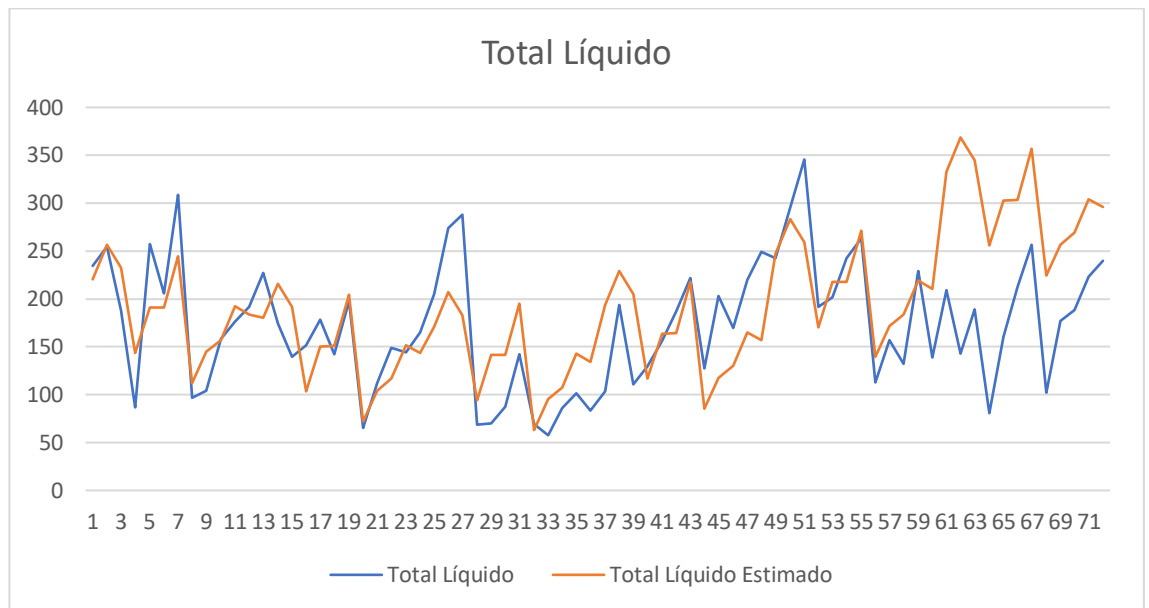
9.2.2.3. Métodos simples - SNaive



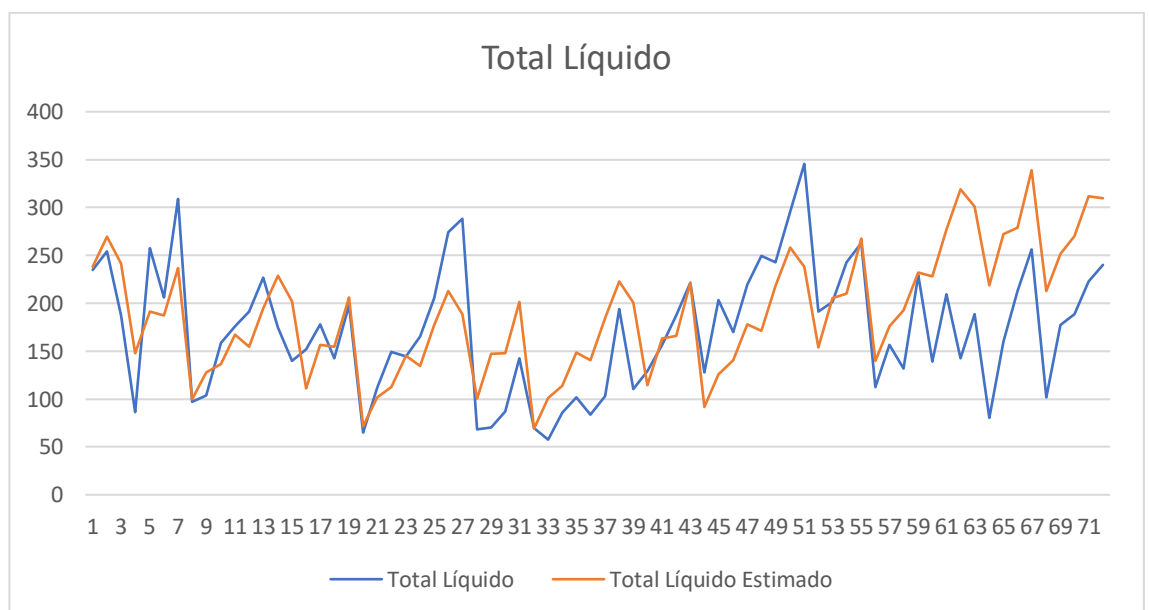
9.2.2.4. Métodos simples - Drift



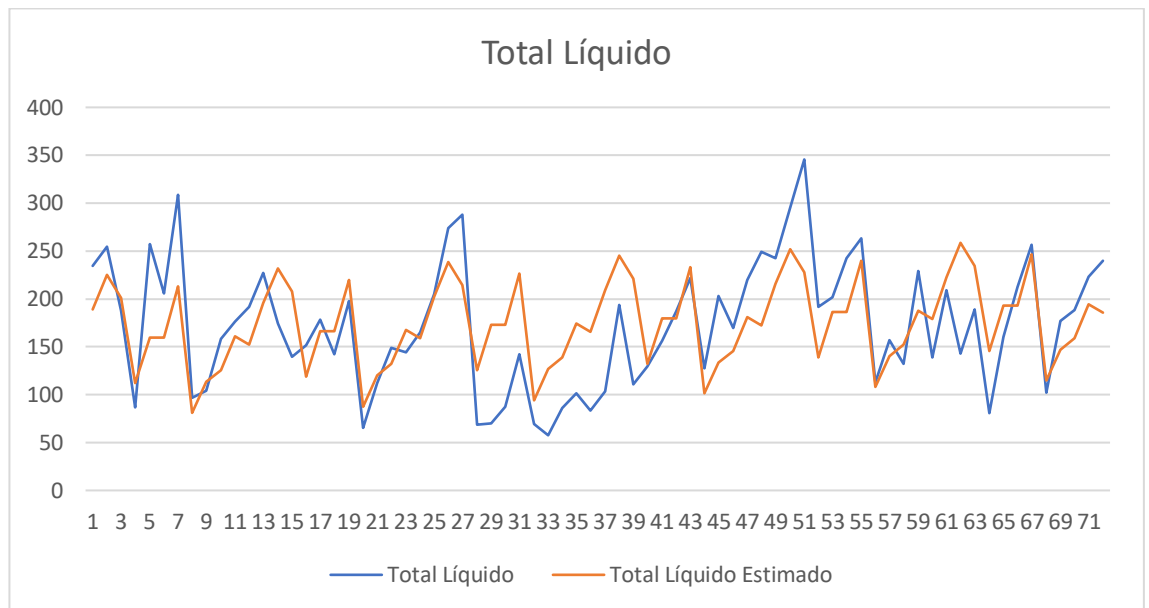
9.2.2.5. Métodos regressivos - Modelo 1



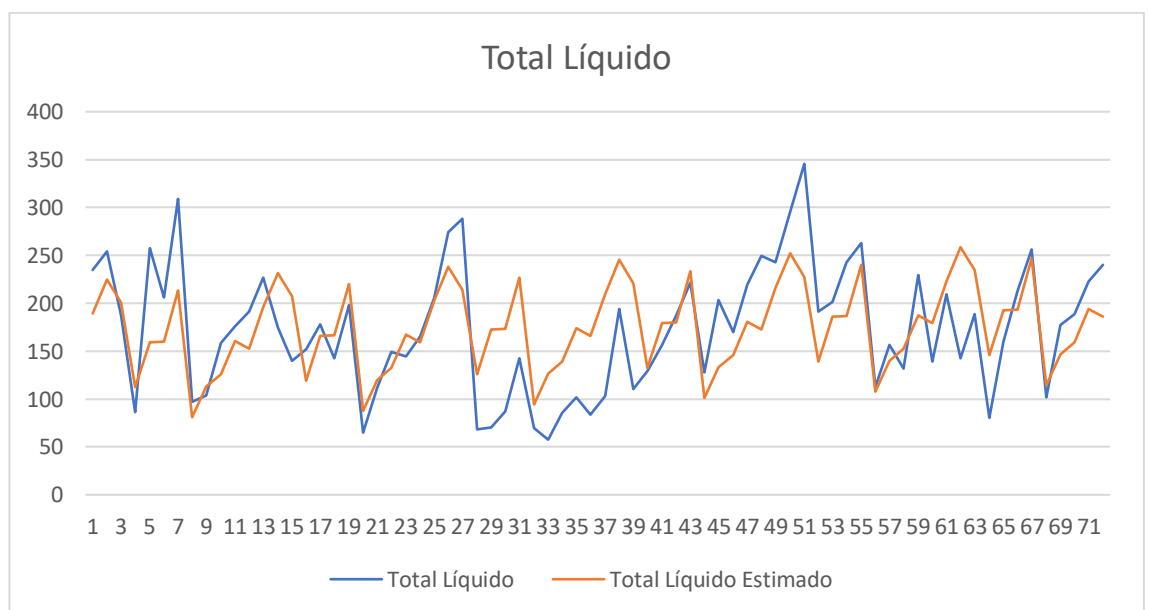
9.2.2.6. Métodos regressivos - Modelo 2



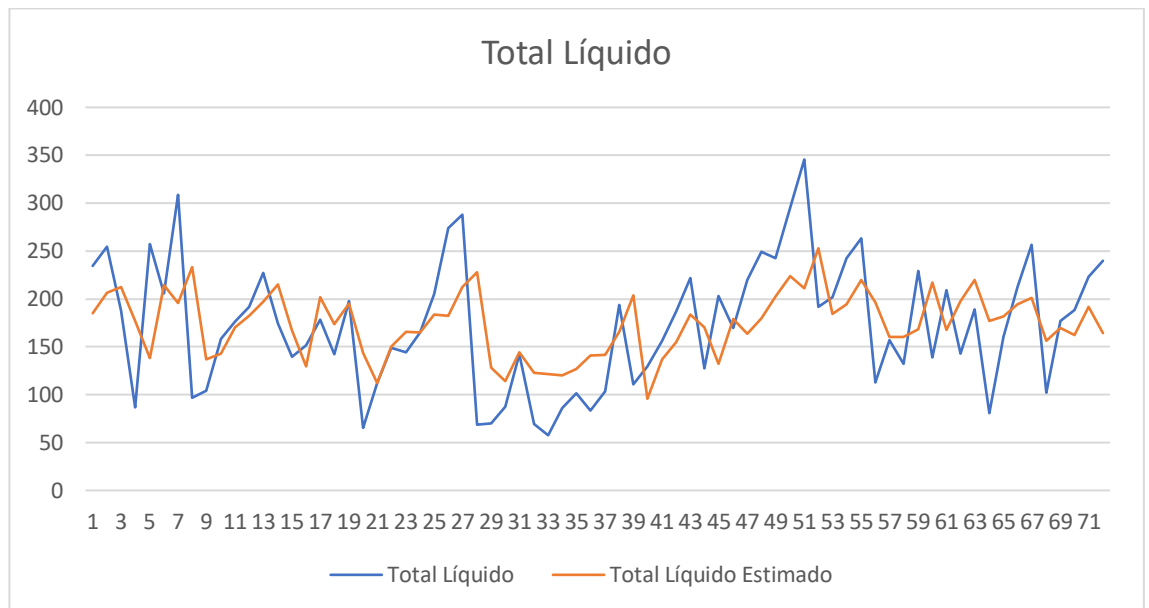
9.2.2.7. Métodos regressivos - Modelo 3



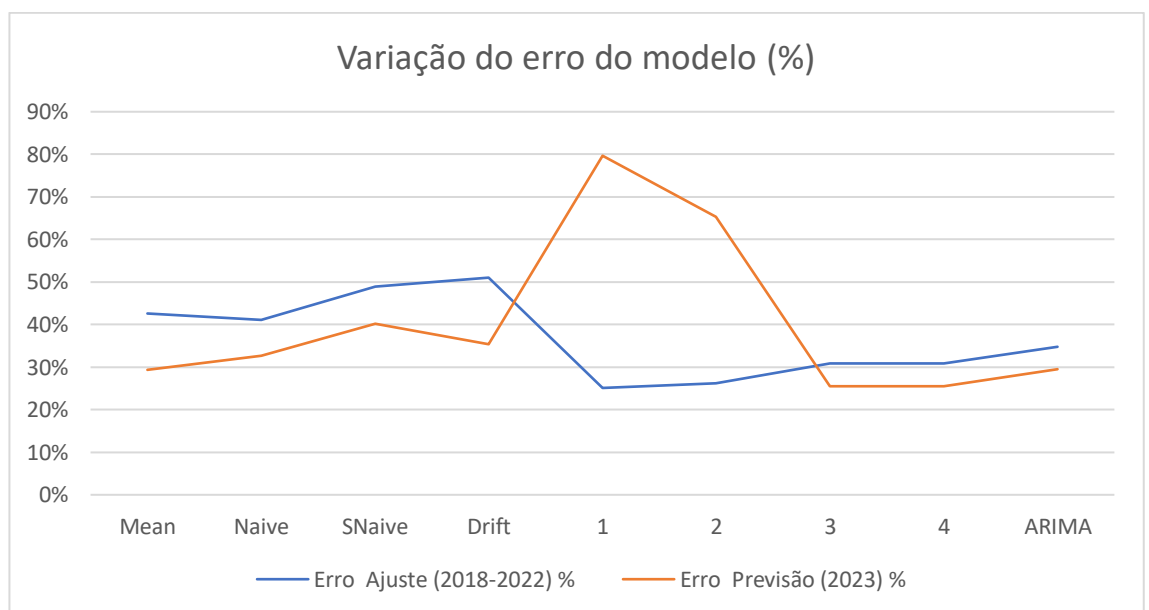
9.2.2.8. Métodos regressivos - Modelo 4



9.2.2.9. ARIMA

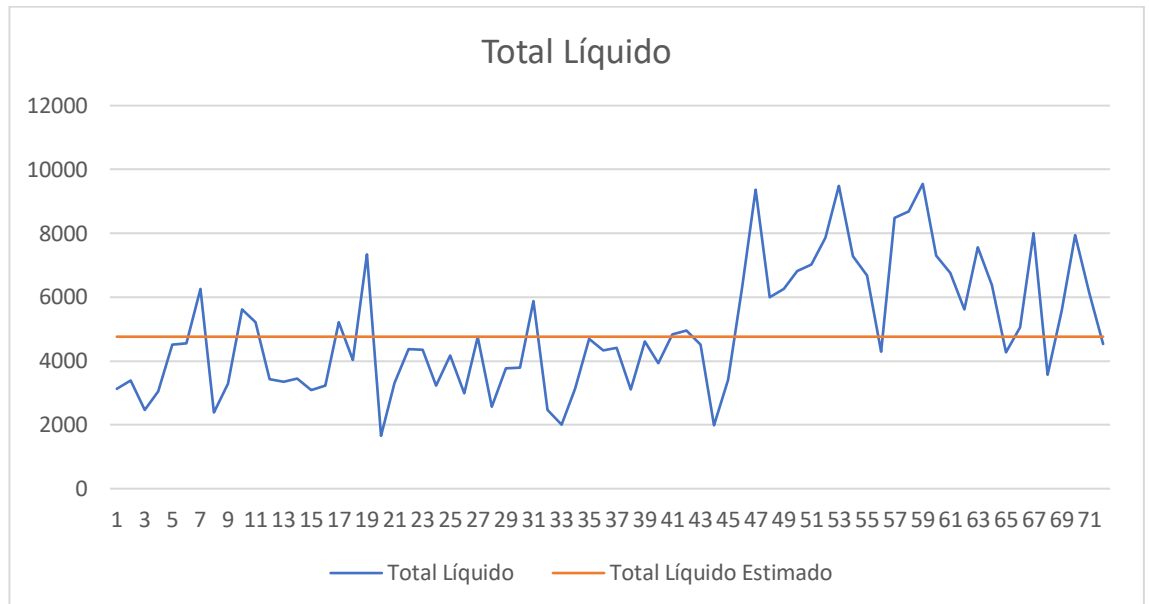


9.2.2.10. Representação gráfica dos erros dos modelos

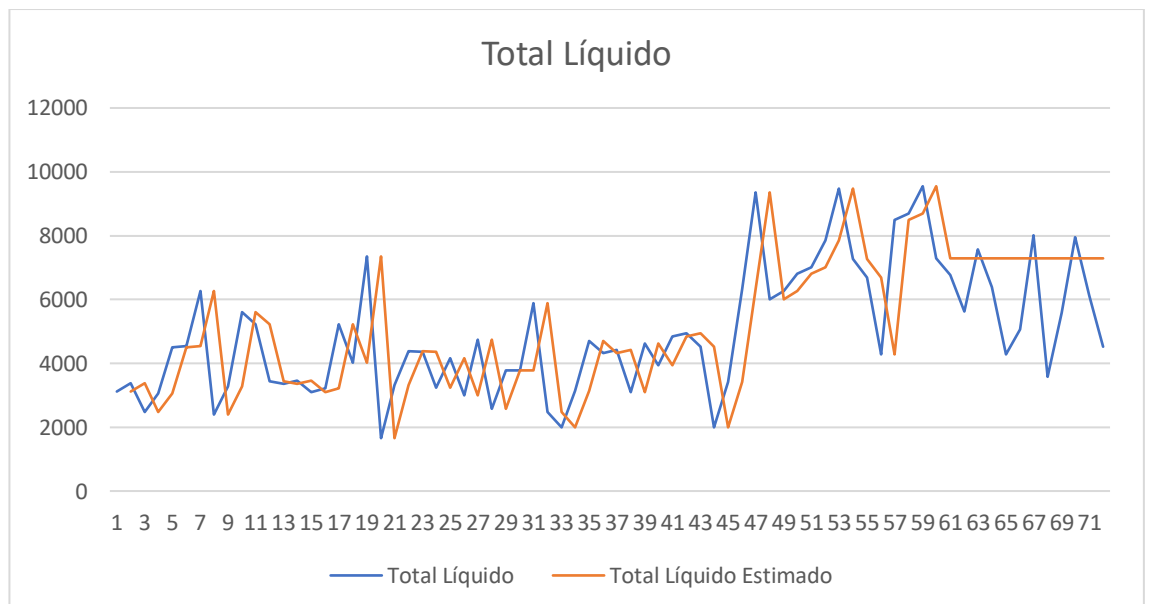


9.2.3. Total Líquido por Cliente – *Cluster 2*

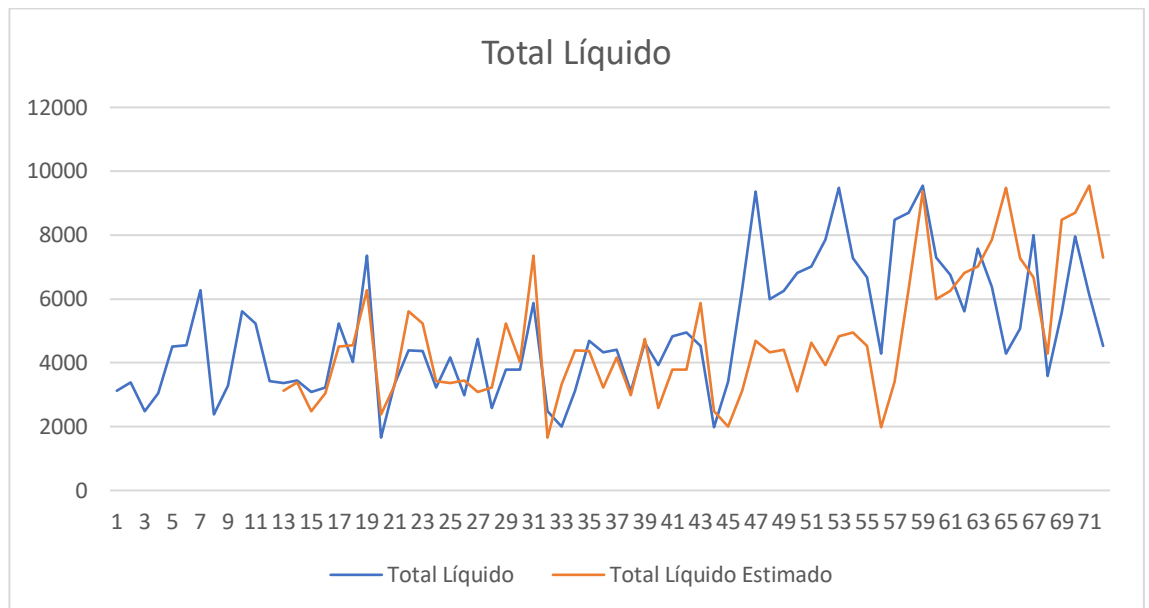
9.2.3.1. Métodos simples - Mean



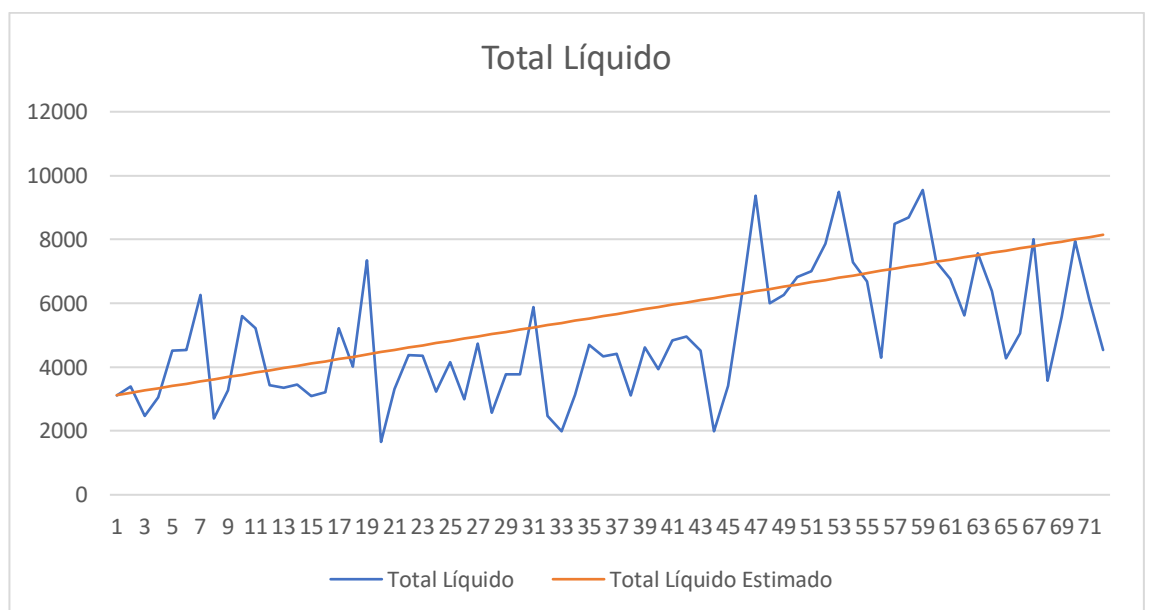
9.2.3.2. Métodos simples - Naive



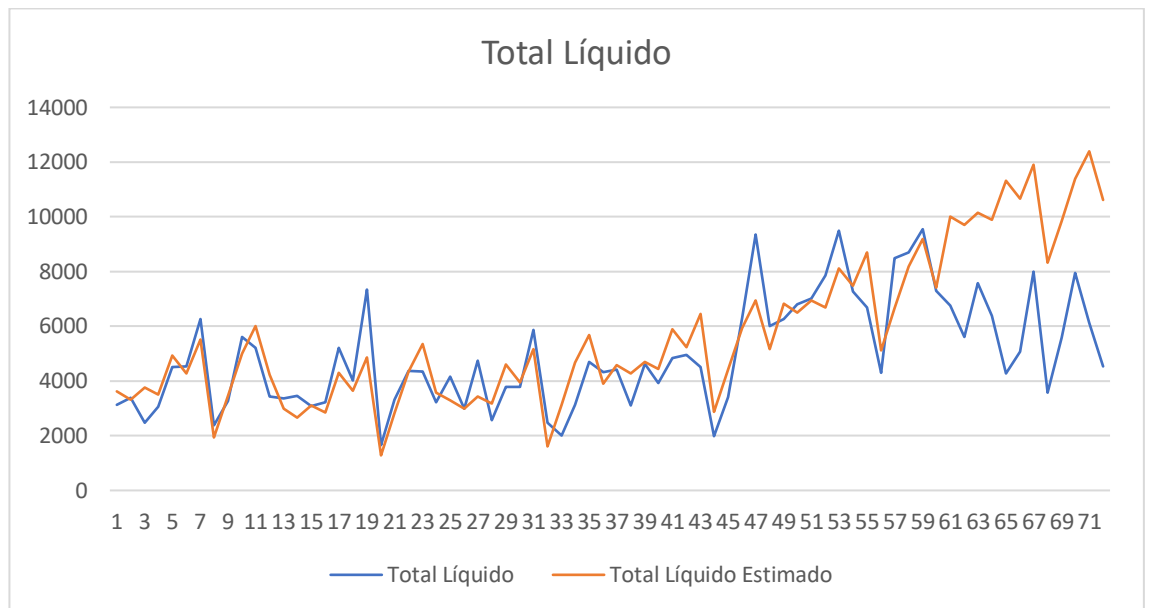
9.2.3.3. Métodos simples - SNaive



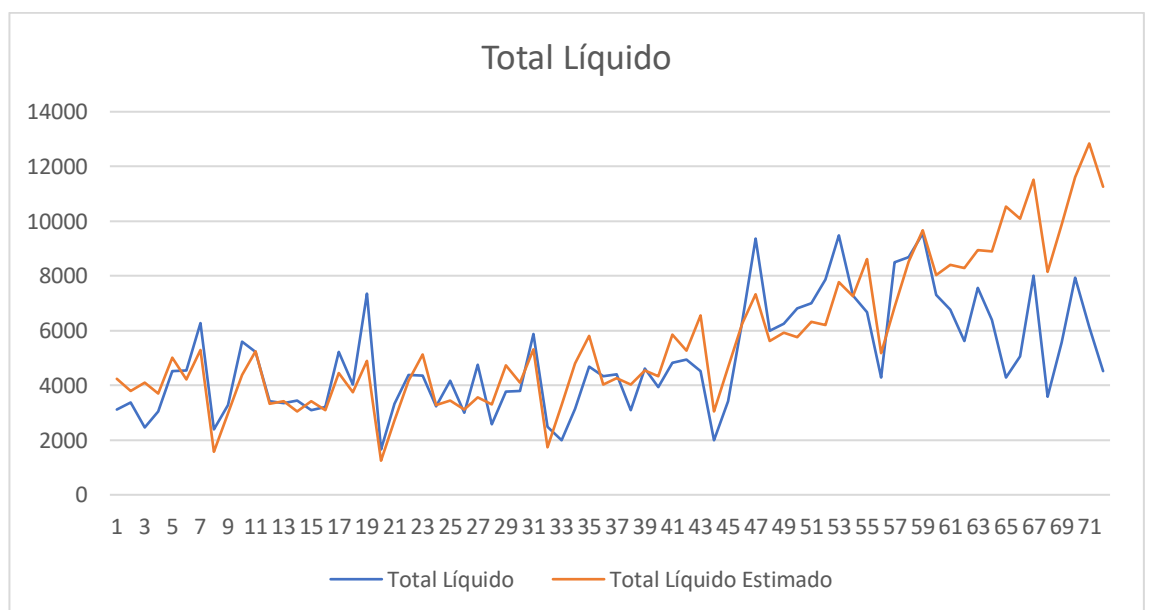
9.2.3.4. Métodos simples - Drift



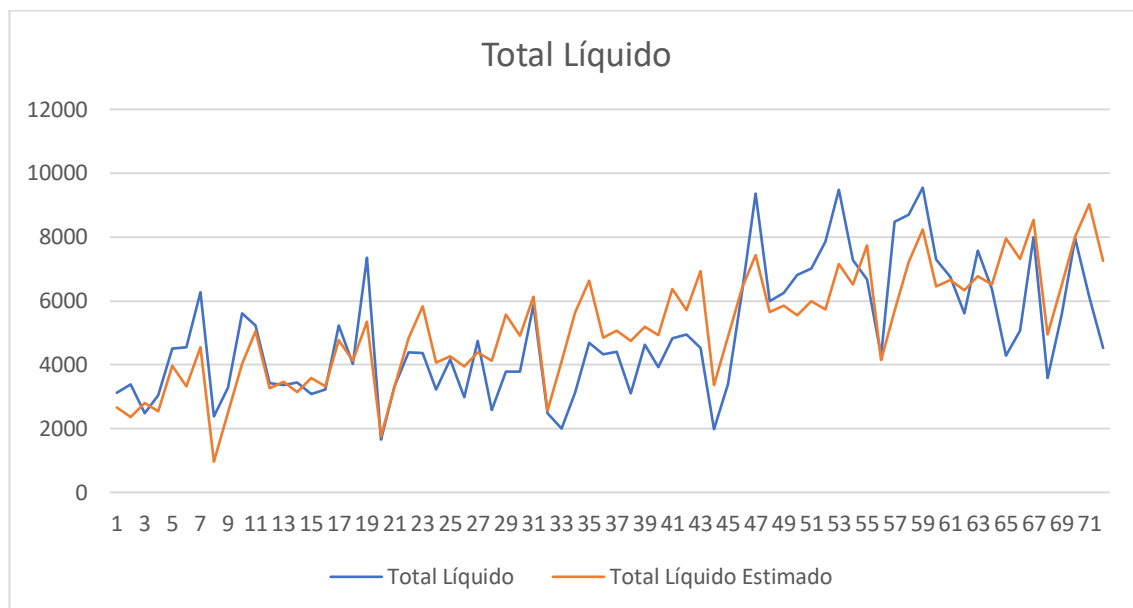
9.2.3.5. Métodos regressivos - Modelo 1



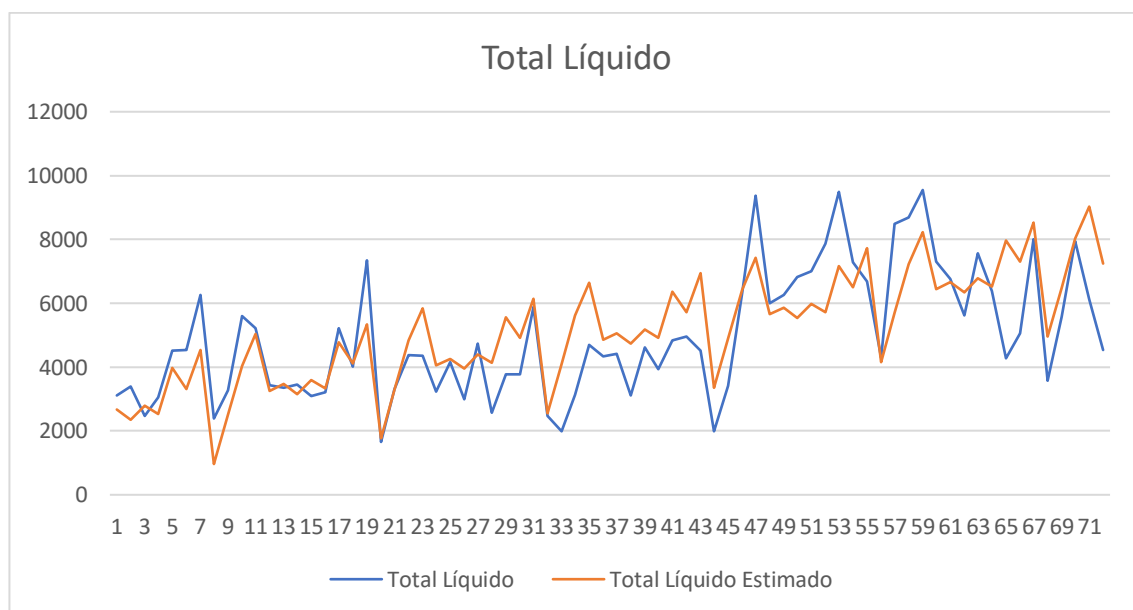
9.2.3.6. Métodos regressivos - Modelo 2



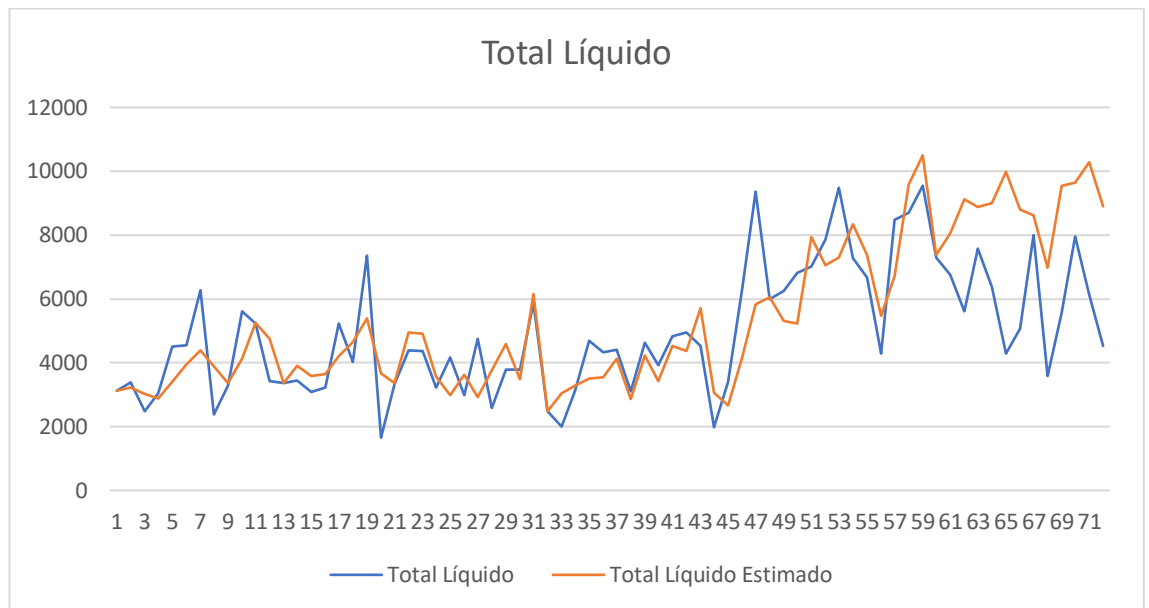
9.2.3.7. Métodos regressivos - Modelo 3



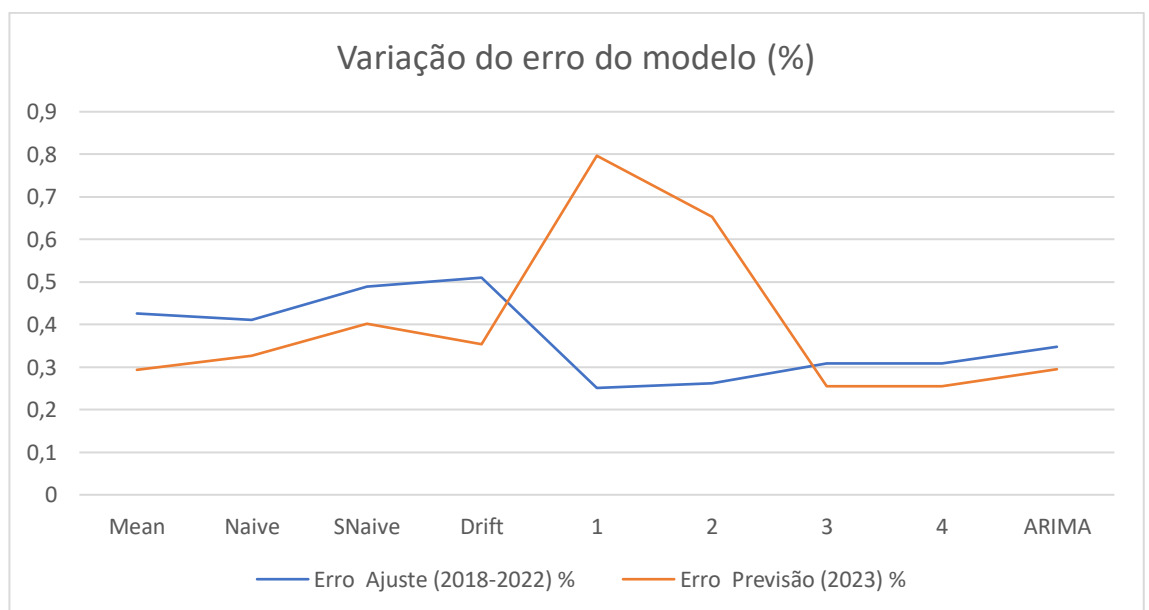
9.2.3.8. Métodos regressivos - Modelo 4



9.2.3.9. ARIMA

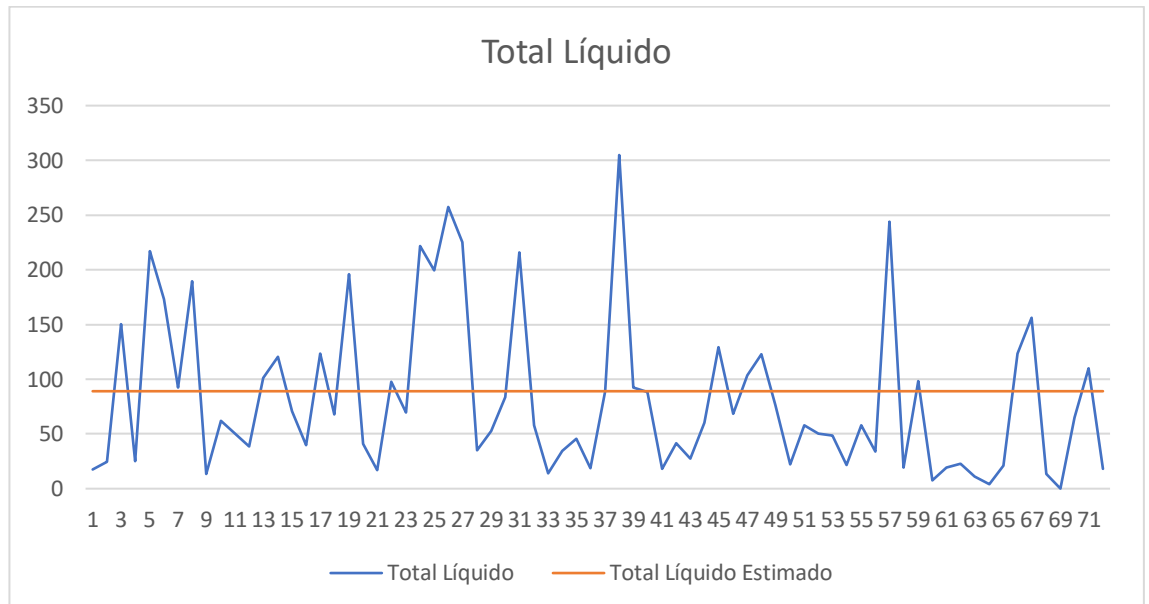


9.2.3.10. Representação gráfica dos erros dos modelos

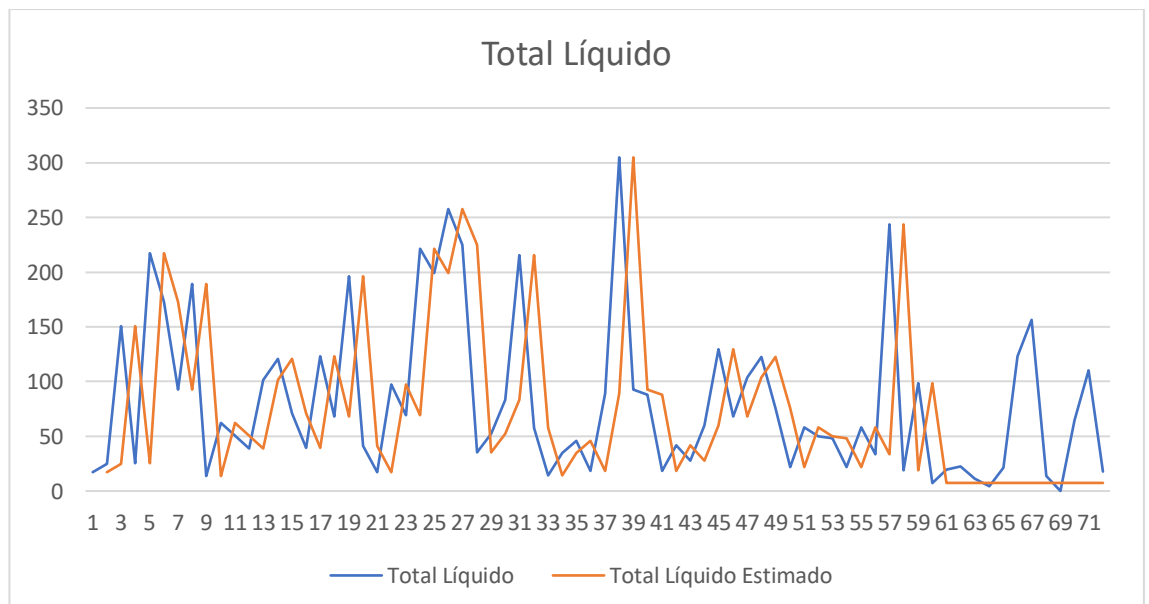


9.2.4. Total Líquido por Tipo de Lavagem – *Cluster 1*

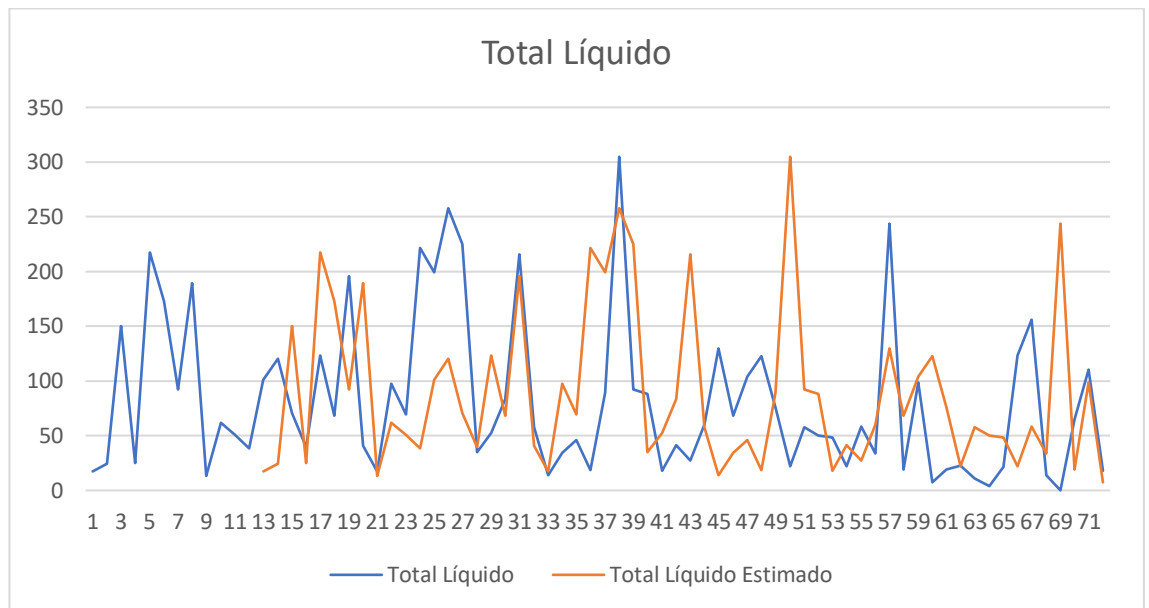
9.2.4.1. Métodos simples - Mean



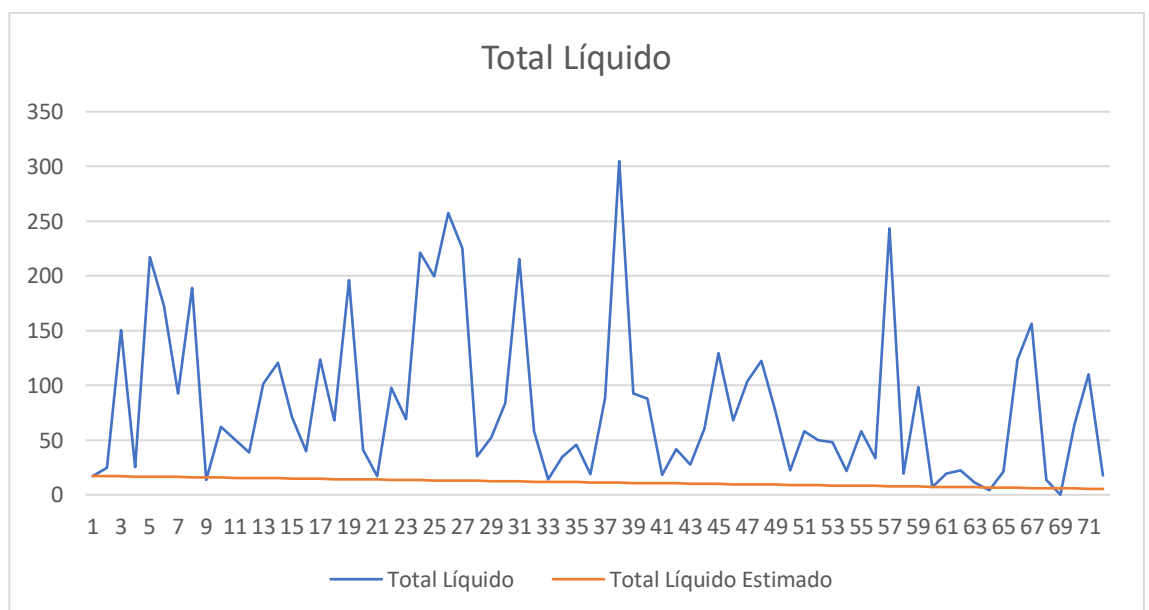
9.2.4.2. Métodos simples - Naive



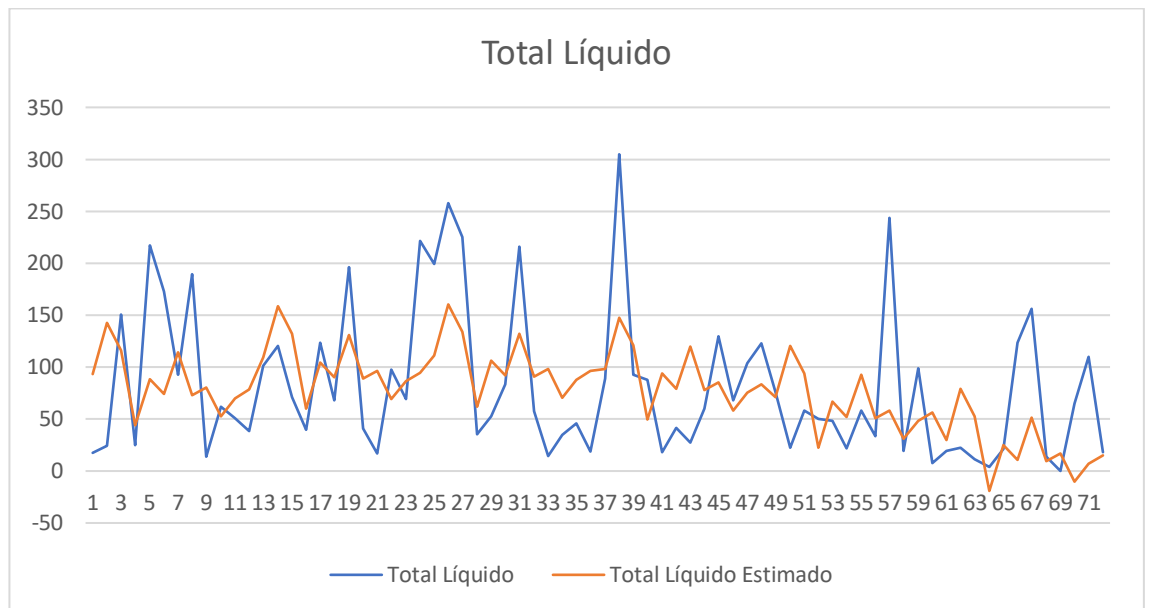
9.2.4.3. Métodos simples - SNaive



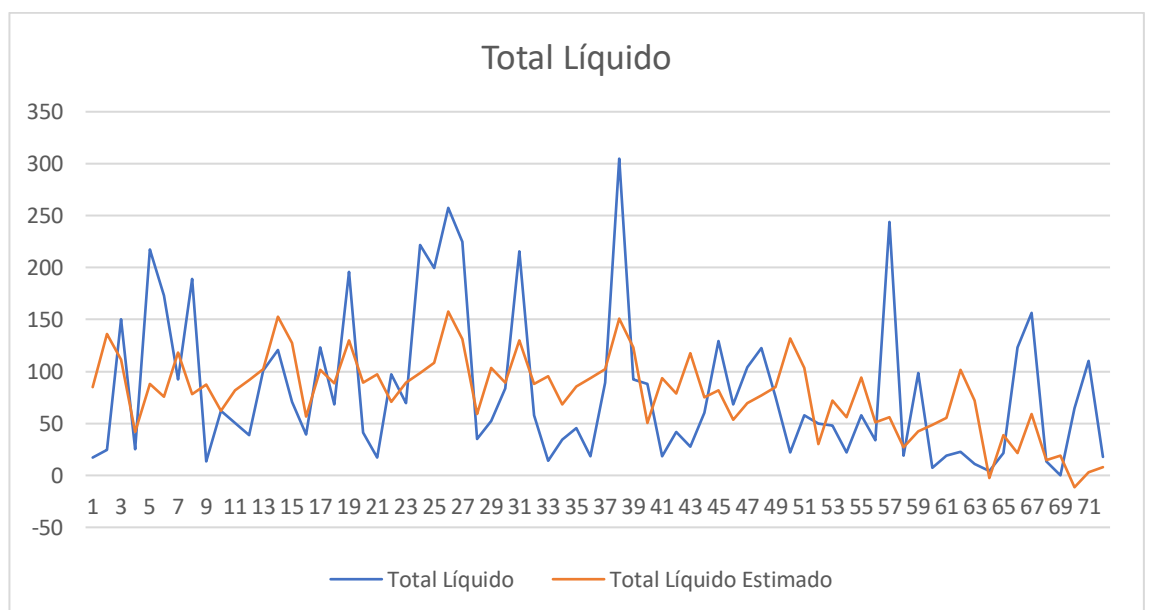
9.2.4.4. Métodos simples - Drift



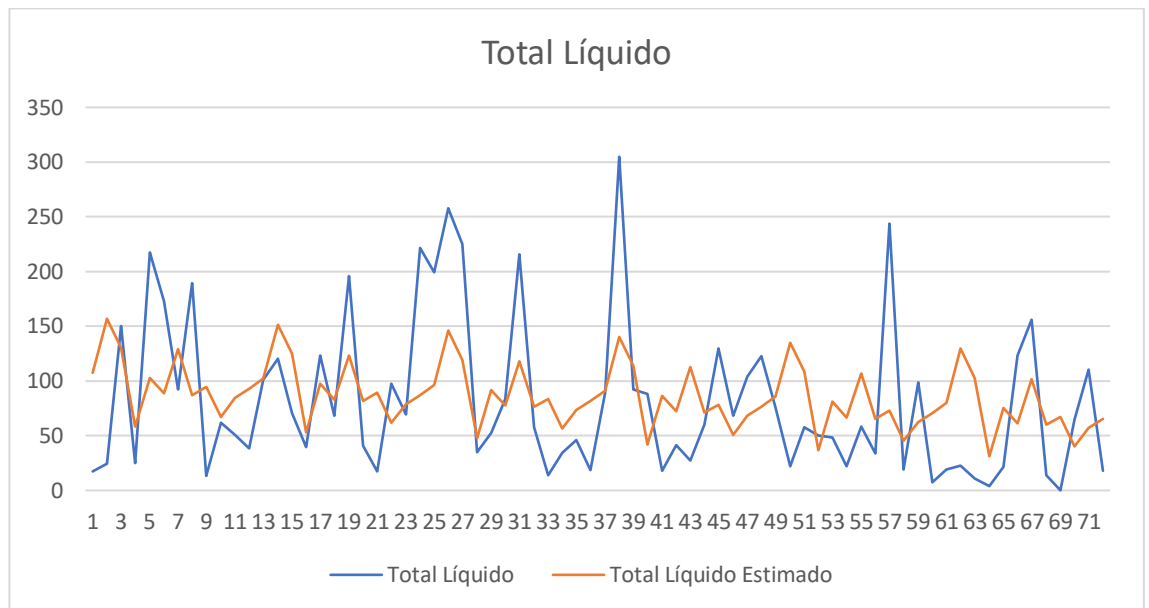
9.2.4.5. Métodos regressivos - Modelo 1



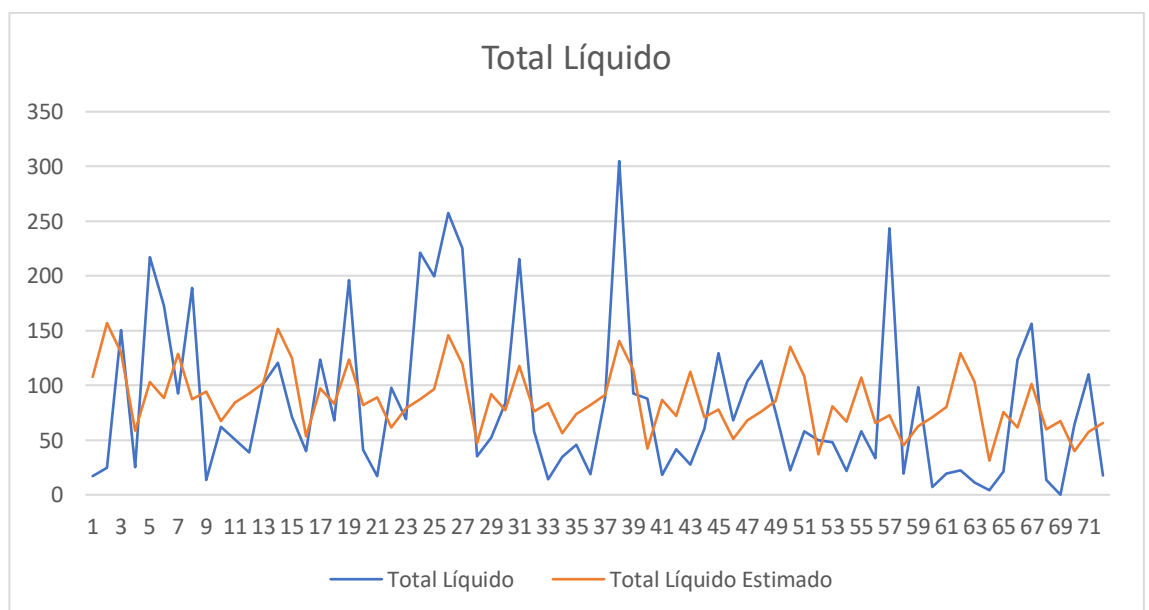
9.2.4.6. Métodos regressivos - Modelo 2



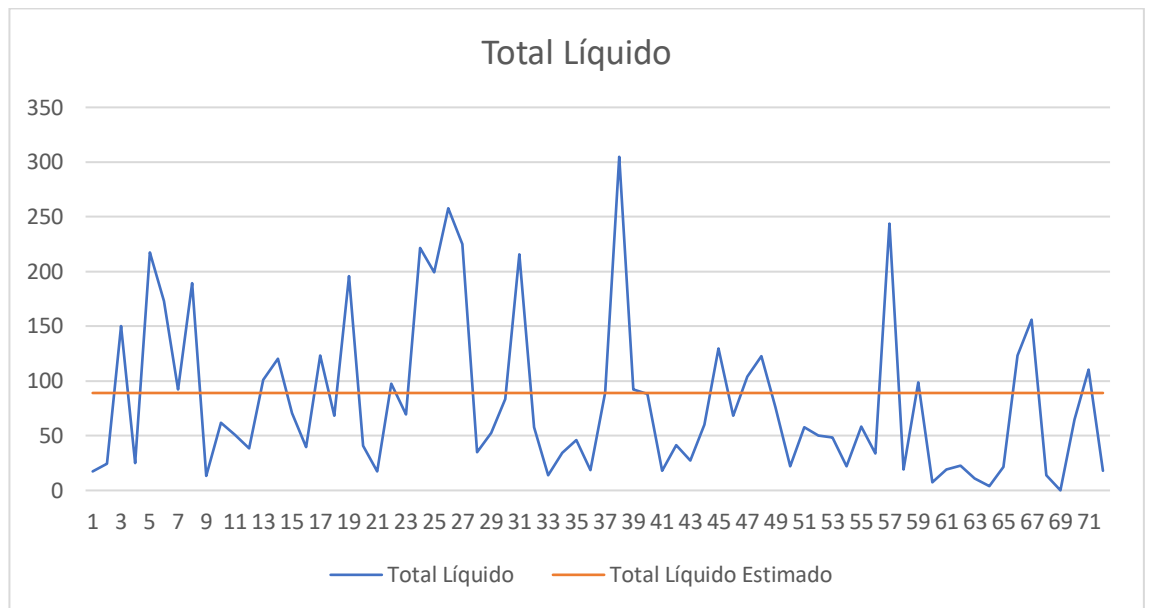
9.2.4.7. Métodos regressivos - Modelo 3



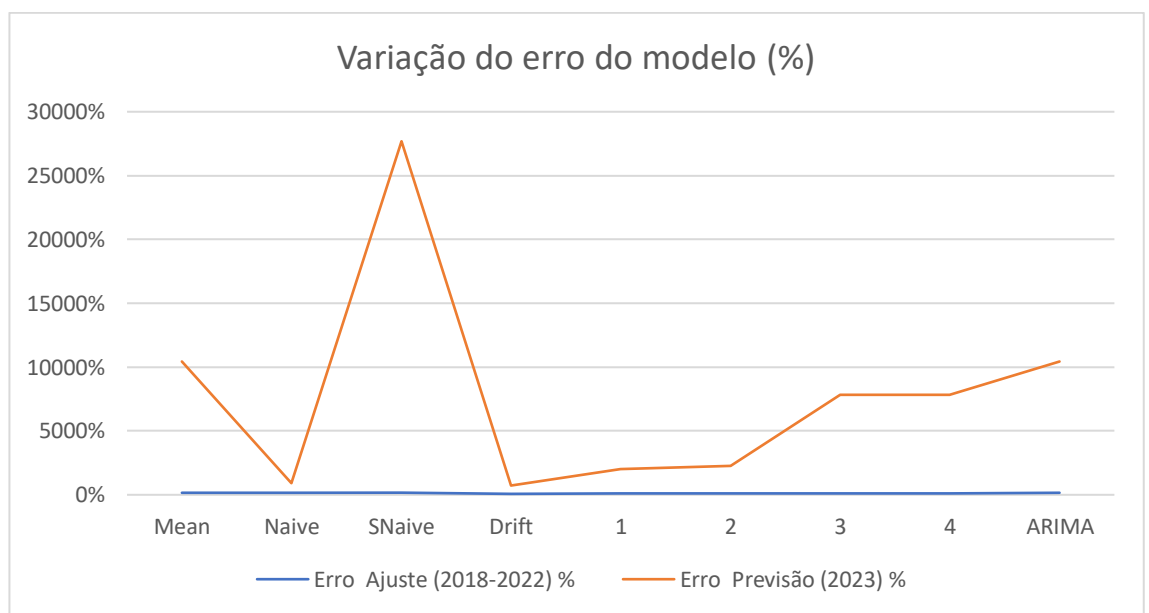
9.2.4.8. Métodos regressivos - Modelo 4



9.2.4.9. ARIMA

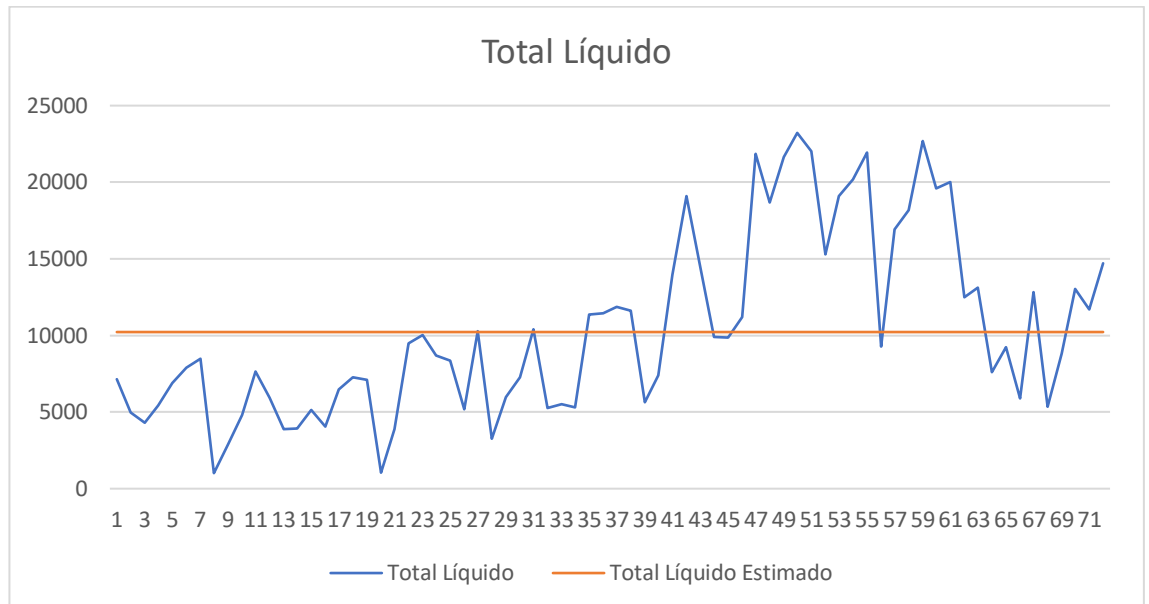


9.2.4.10. Representação gráfica dos erros dos modelos

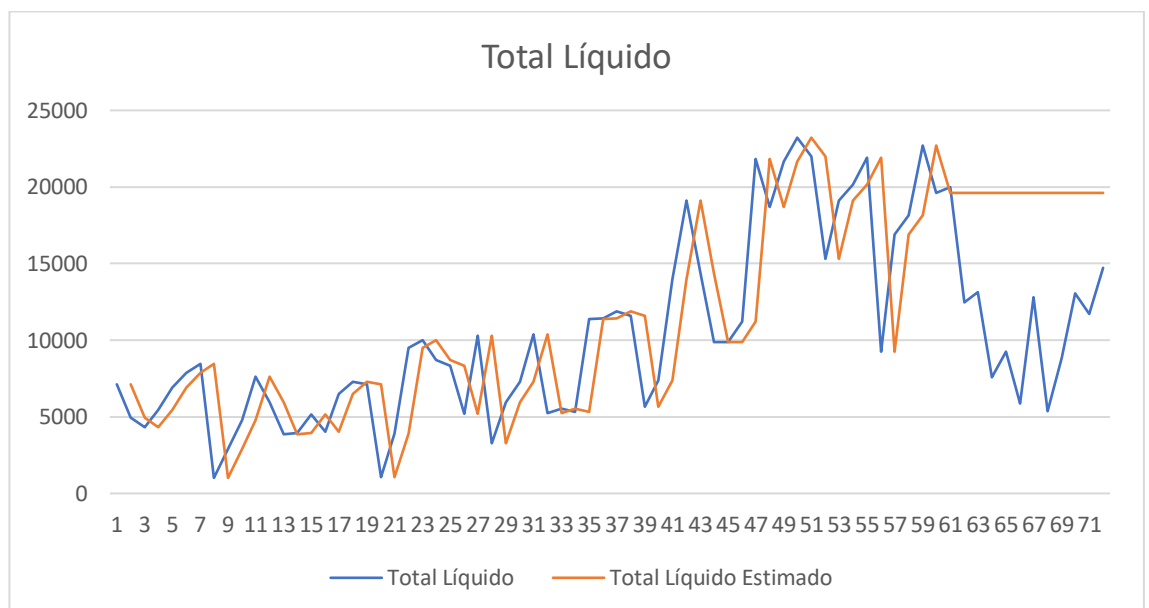


9.2.5. Total Líquido por Tipo de Lavagem – *Cluster 2*

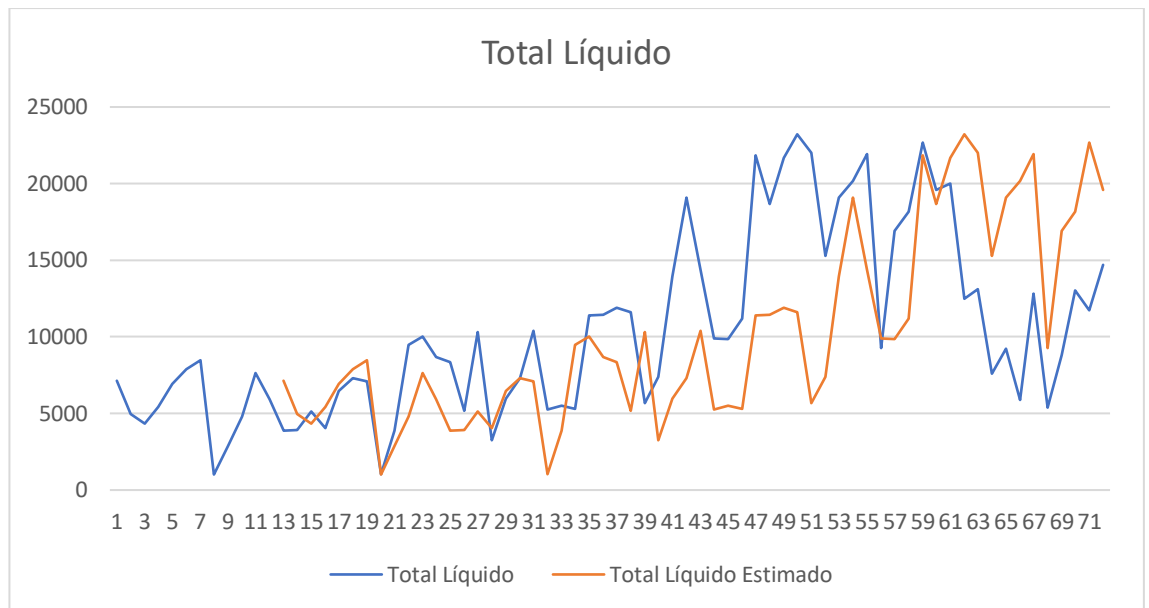
9.2.5.1. Métodos simples - Mean



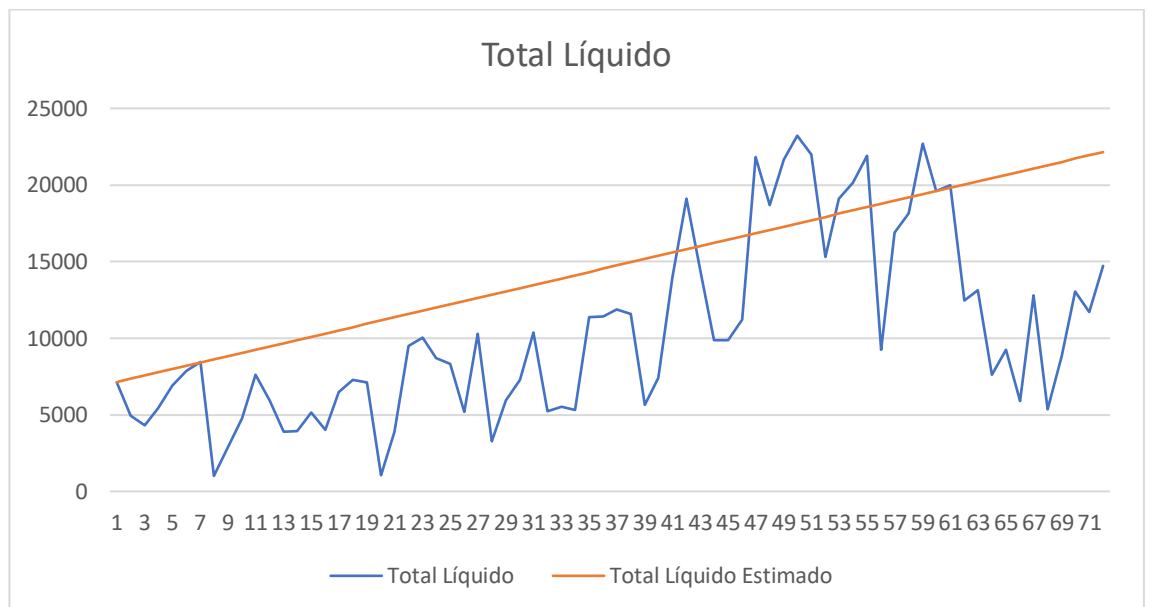
9.2.5.2. Métodos simples - Naive



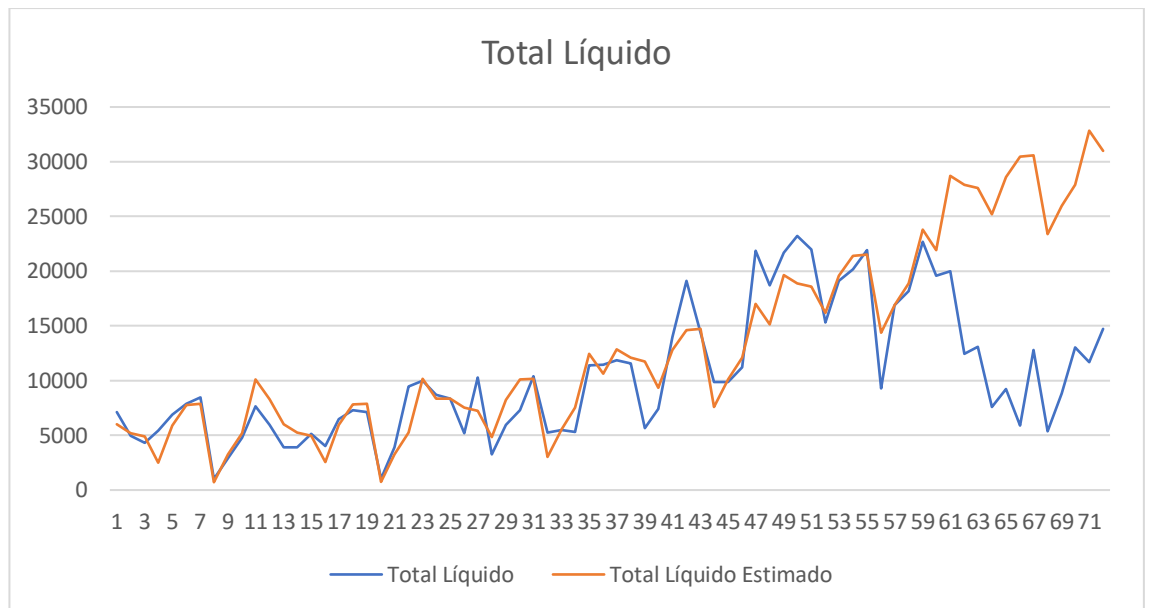
9.2.5.3. Métodos simples - SNaive



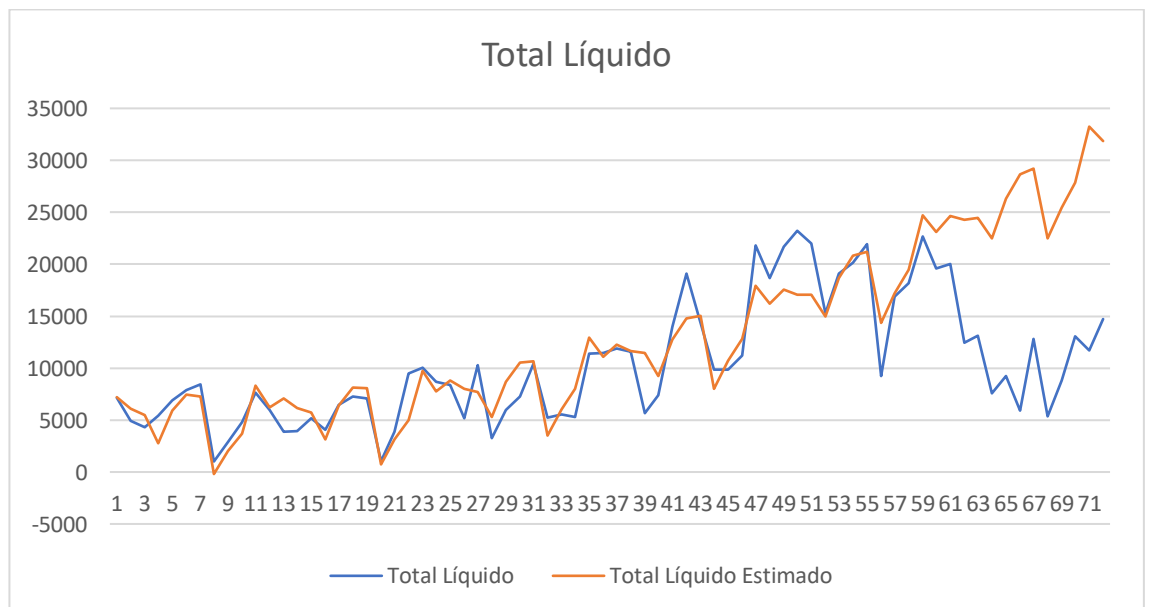
9.2.5.4. Métodos simples - Drift



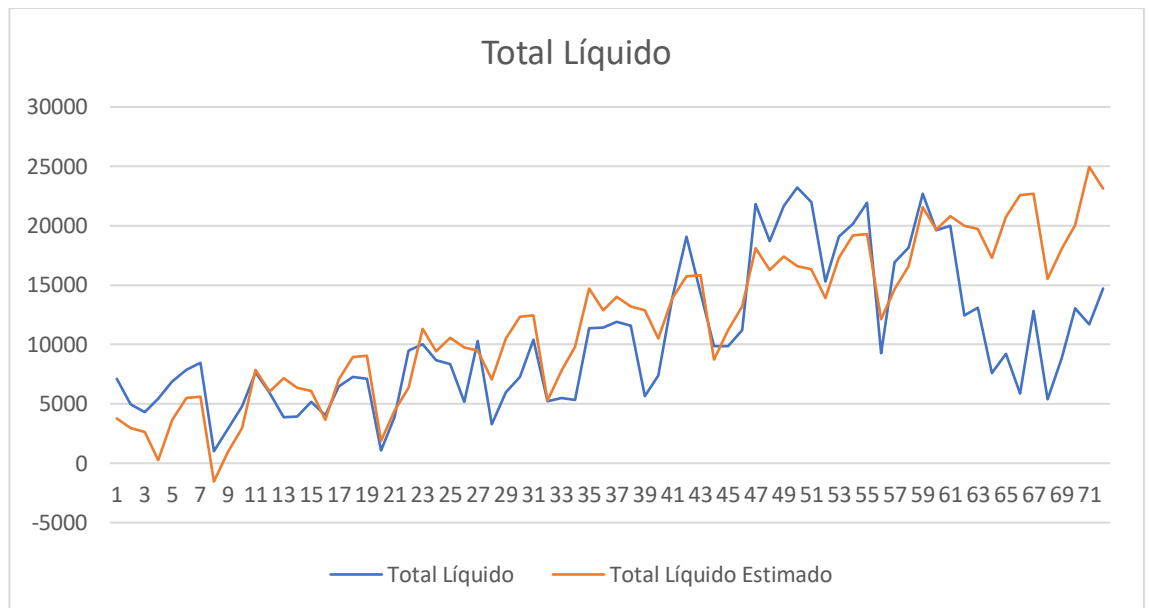
9.2.5.5. Métodos regressivos - Modelo 1



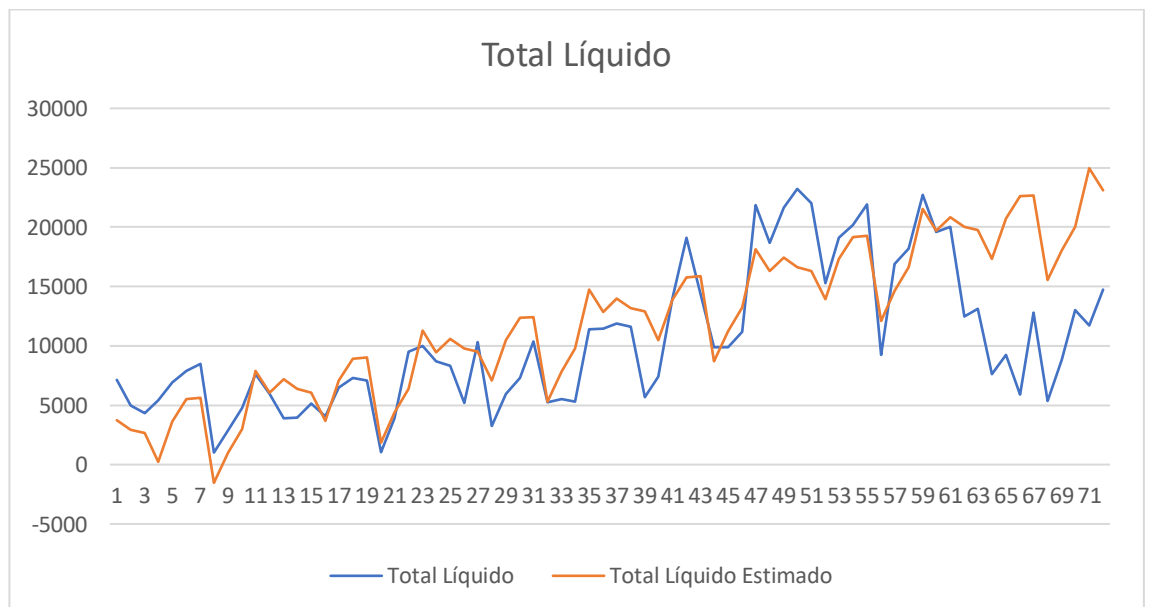
9.2.5.6. Métodos regressivos - Modelo 2



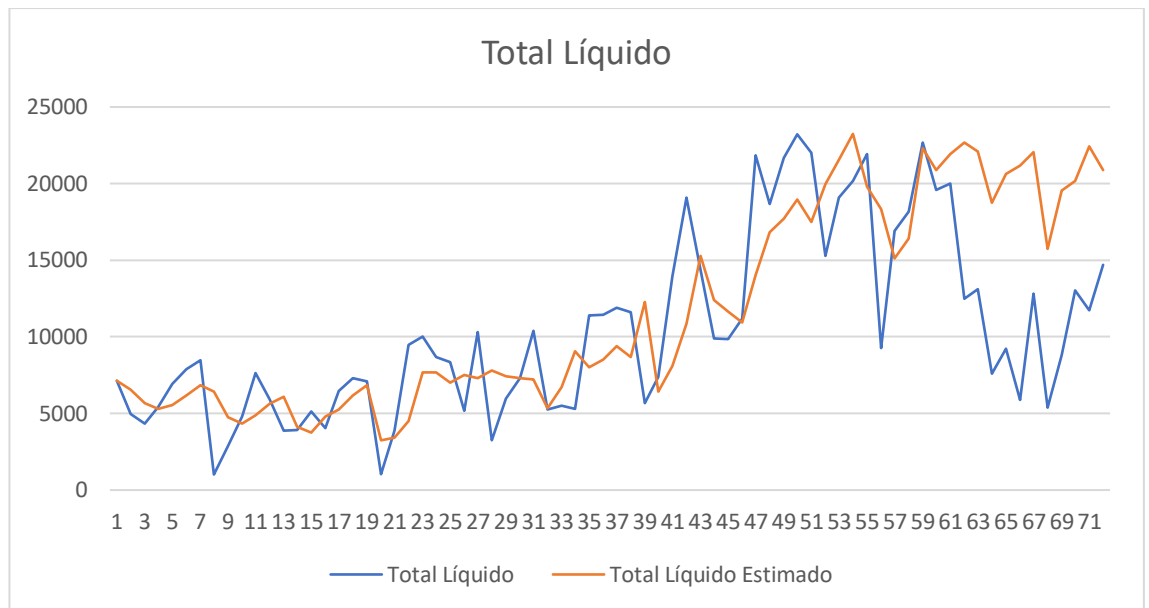
9.2.5.7. Métodos regressivos - Modelo 3



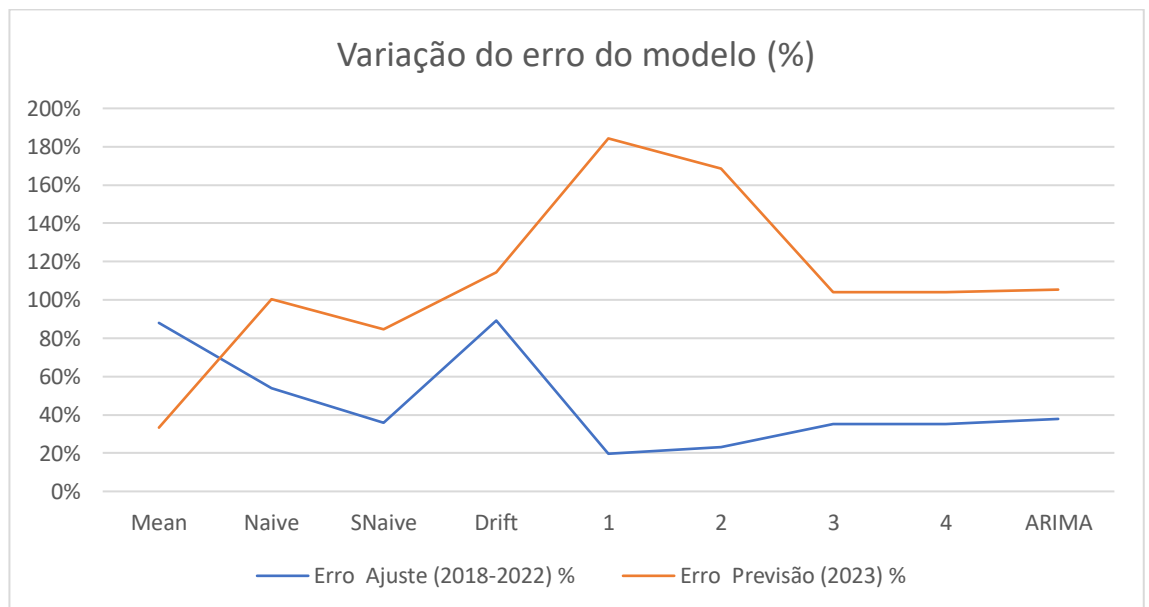
9.2.5.8. Métodos regressivos - Modelo 4



9.2.5.9. ARIMA

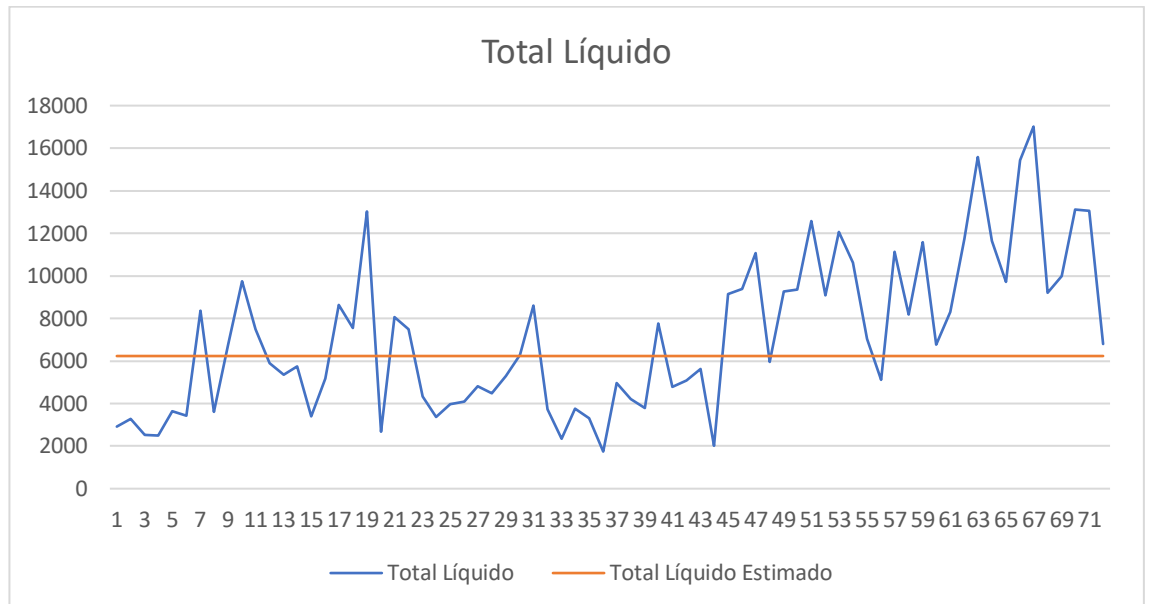


9.2.5.10. Representação gráfica dos erros dos modelos

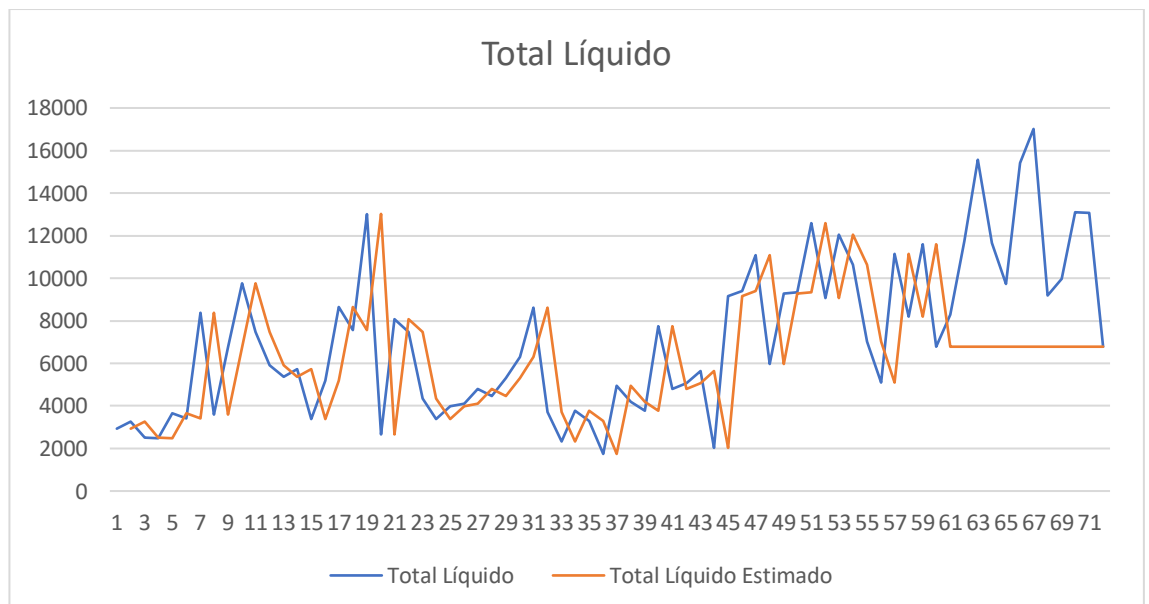


9.2.6. Total Líquido por Tipo de Lavagem – *Cluster 3*

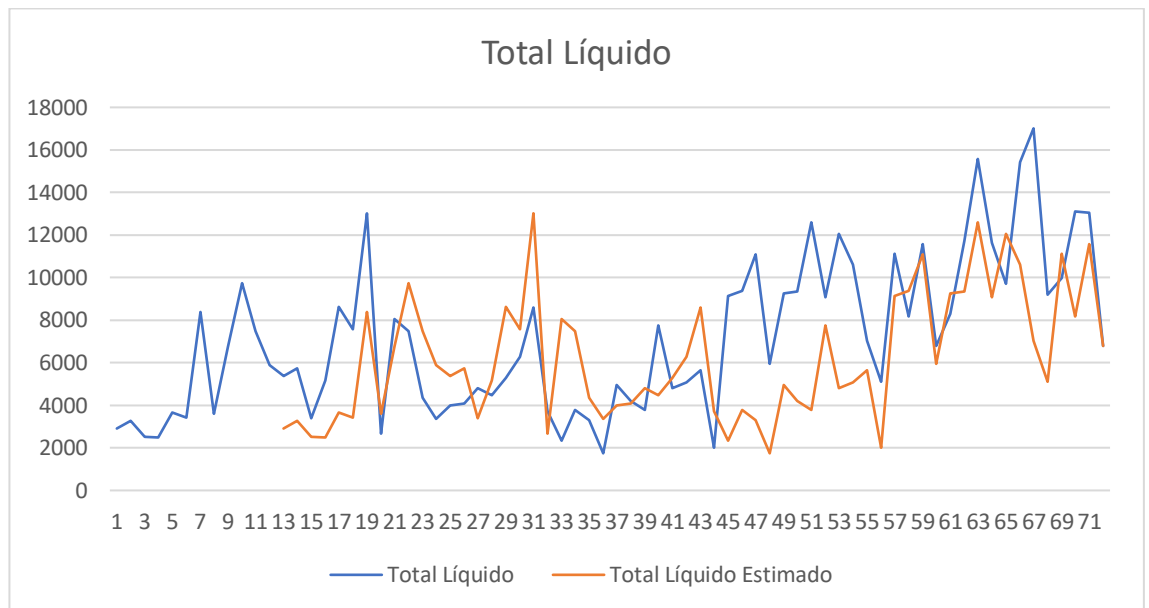
9.2.6.1. Métodos simples - Mean



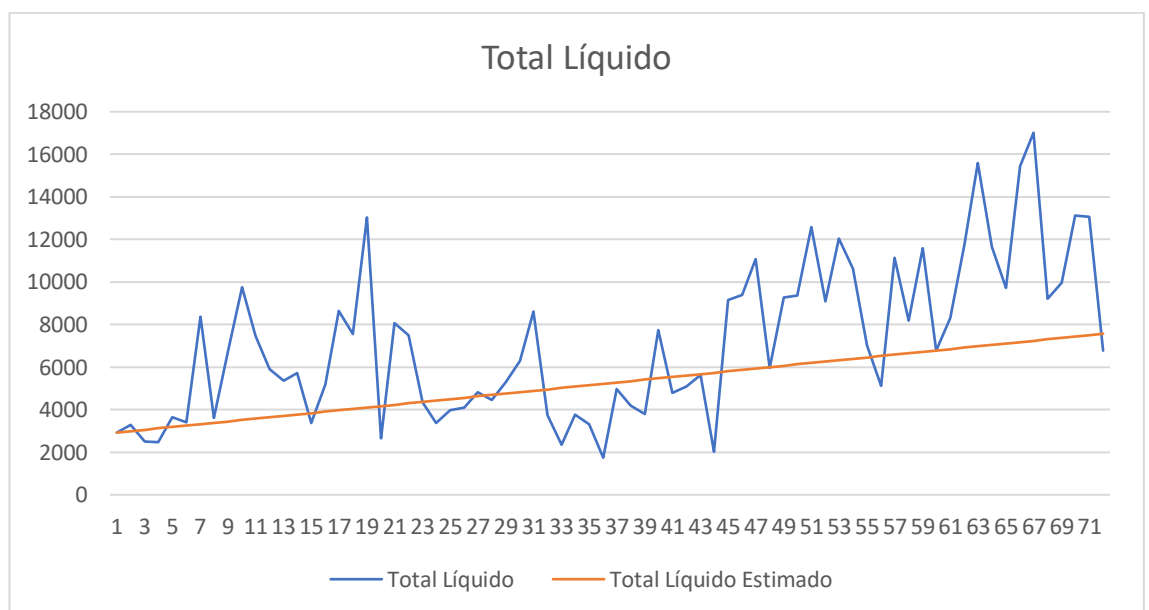
9.2.6.2. Métodos simples - Naive



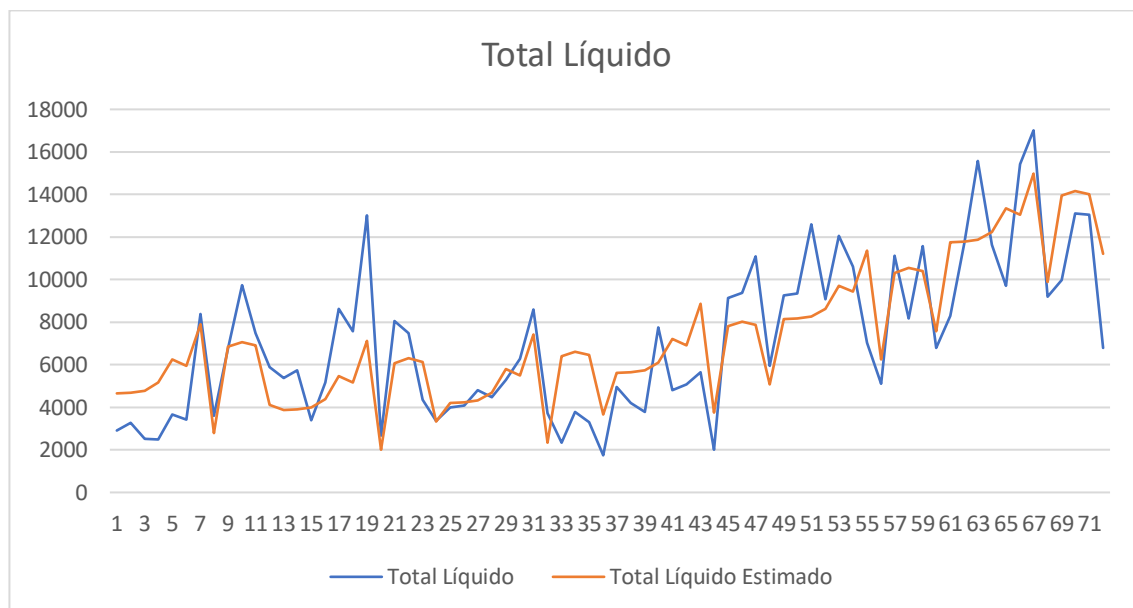
9.2.6.3. Métodos simples - SNaive



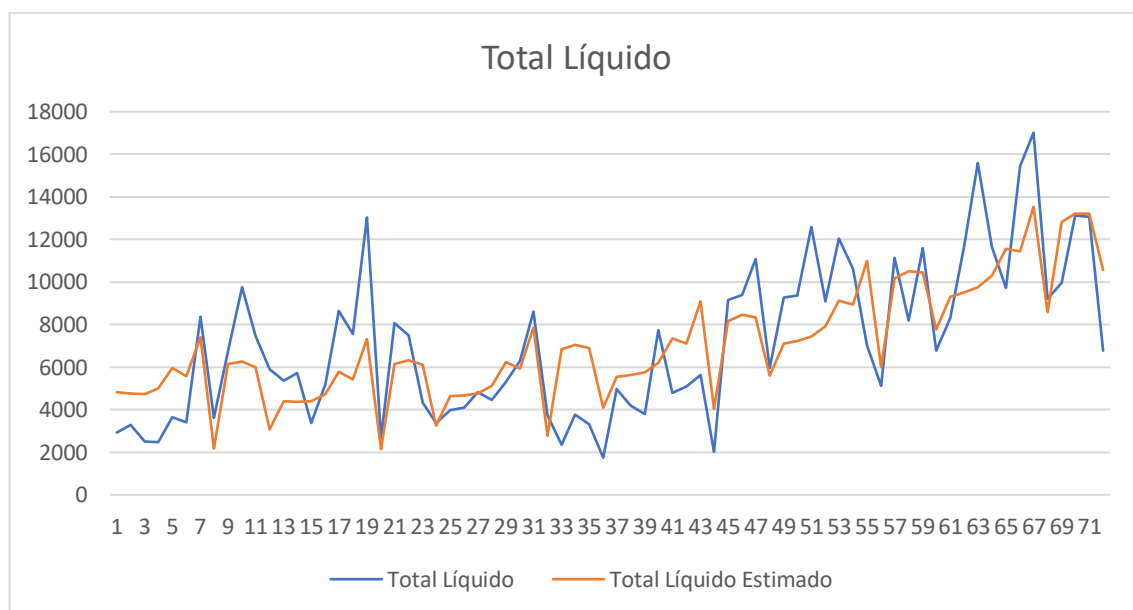
9.2.6.4. Métodos simples - Drift



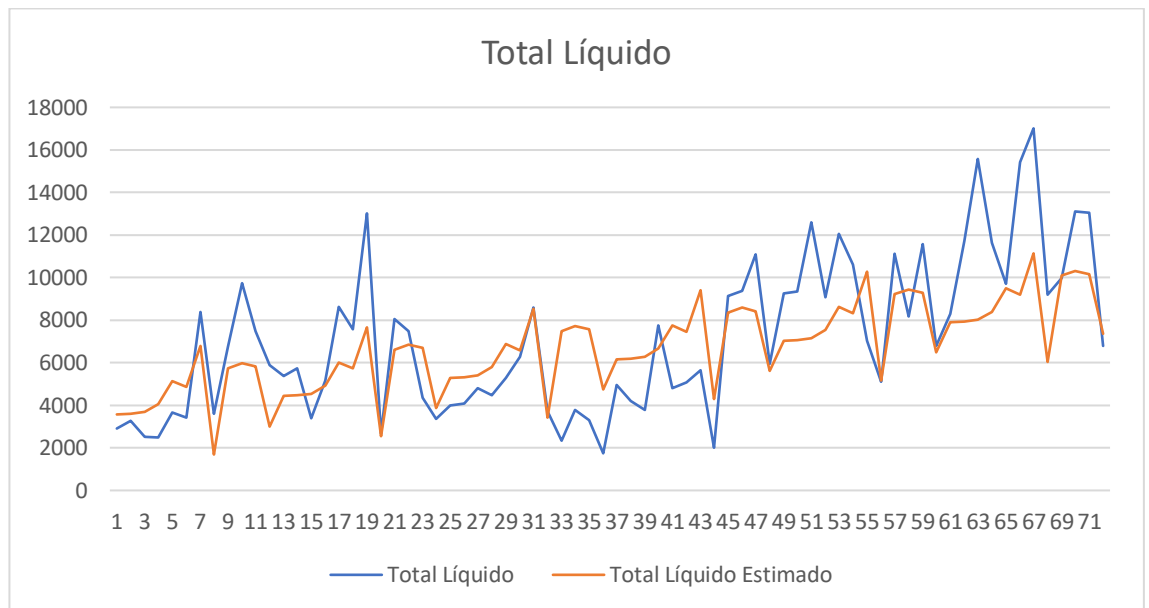
9.2.6.5. Métodos regressivos - Modelo 1



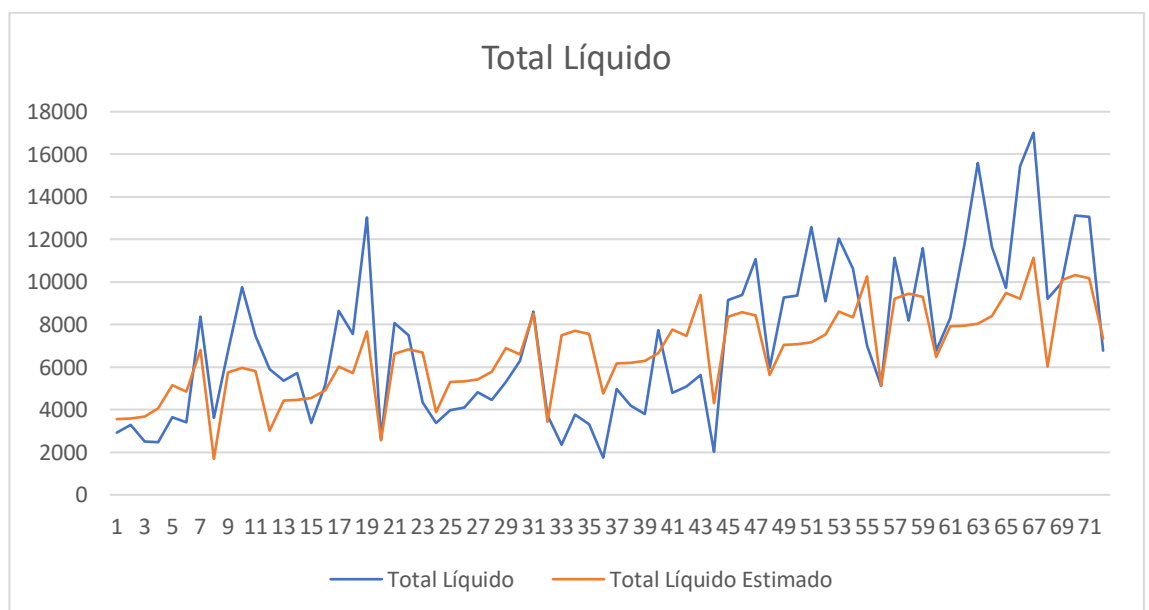
9.2.6.6. Métodos regressivos - Modelo 2



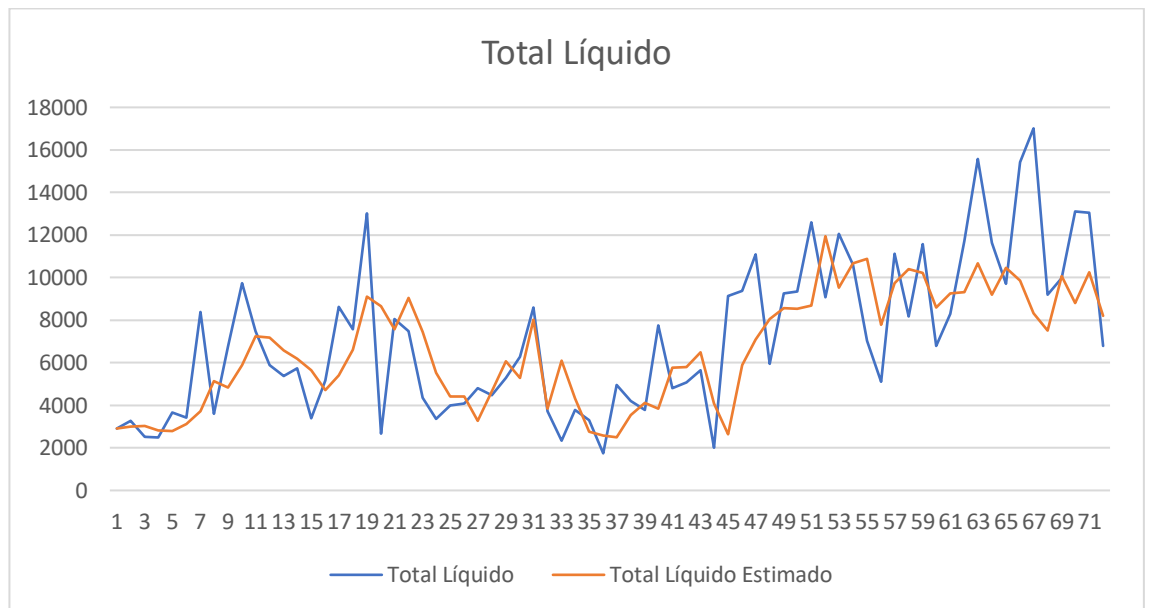
9.2.6.7. Métodos regressivos - Modelo 3



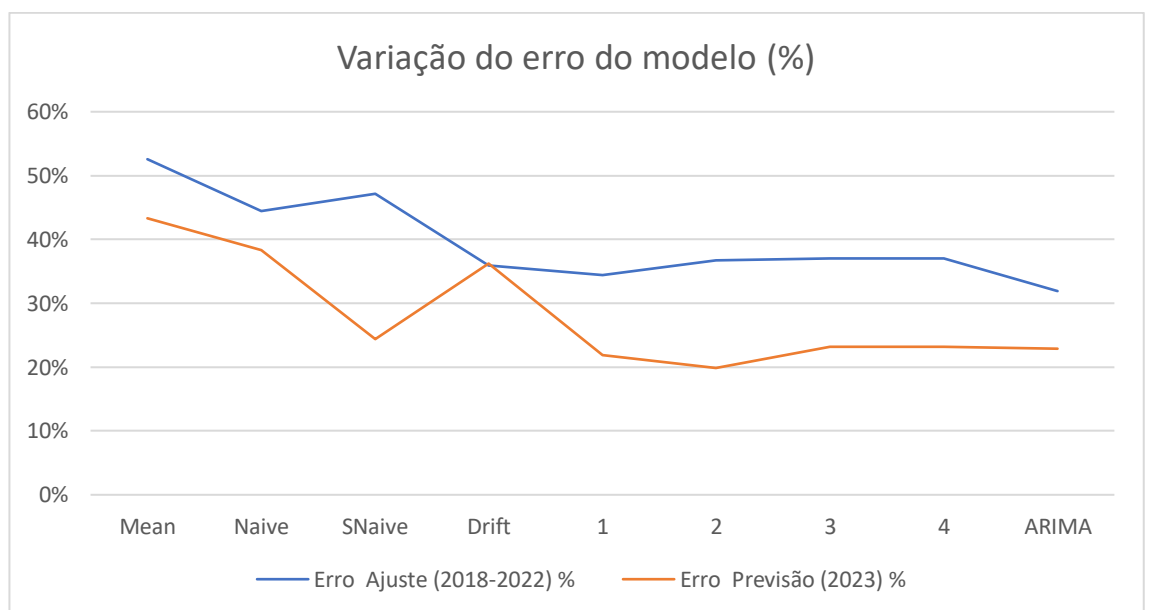
9.2.6.8. Métodos regressivos - Modelo 4



9.2.6.9. ARIMA

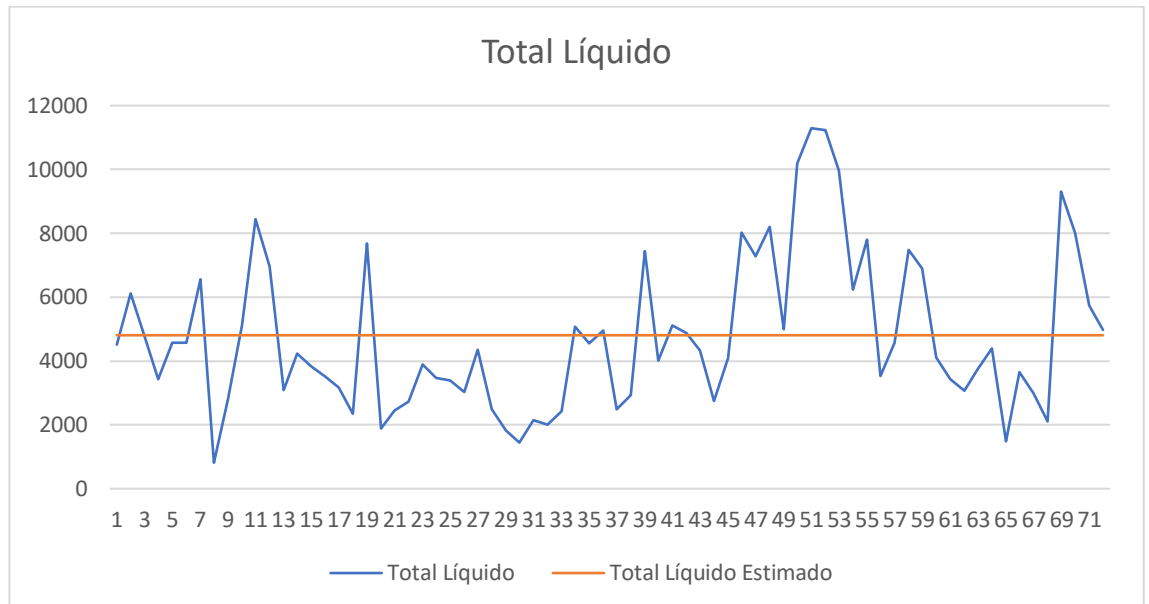


9.2.6.10. Representação gráfica dos erros dos modelos

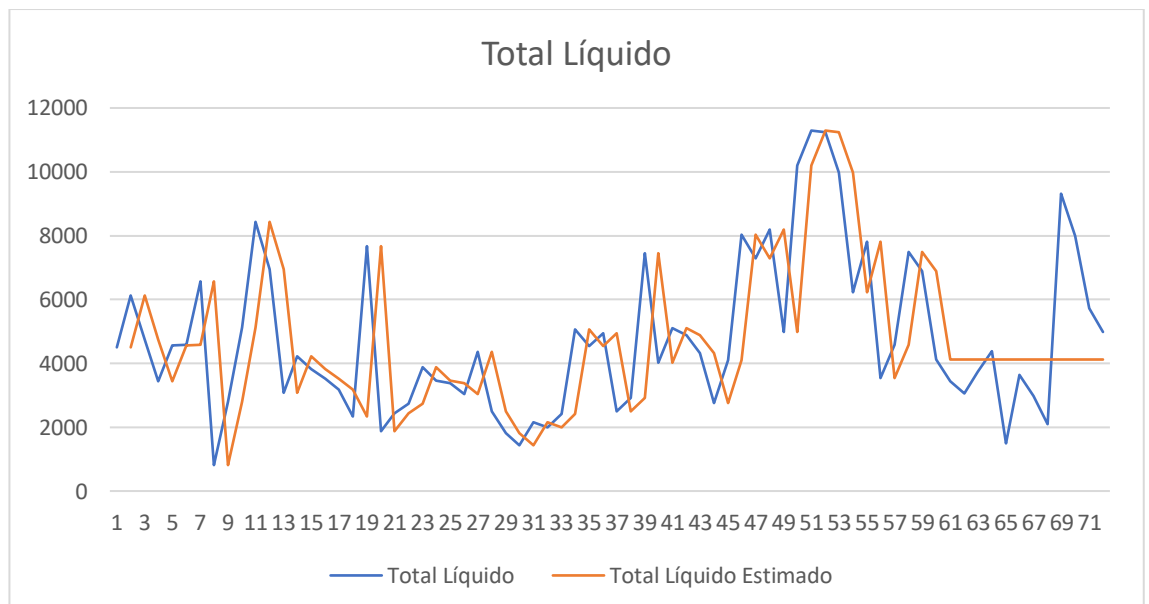


9.2.7. Total Líquido por Tipo de Lavagem – *Cluster 4*

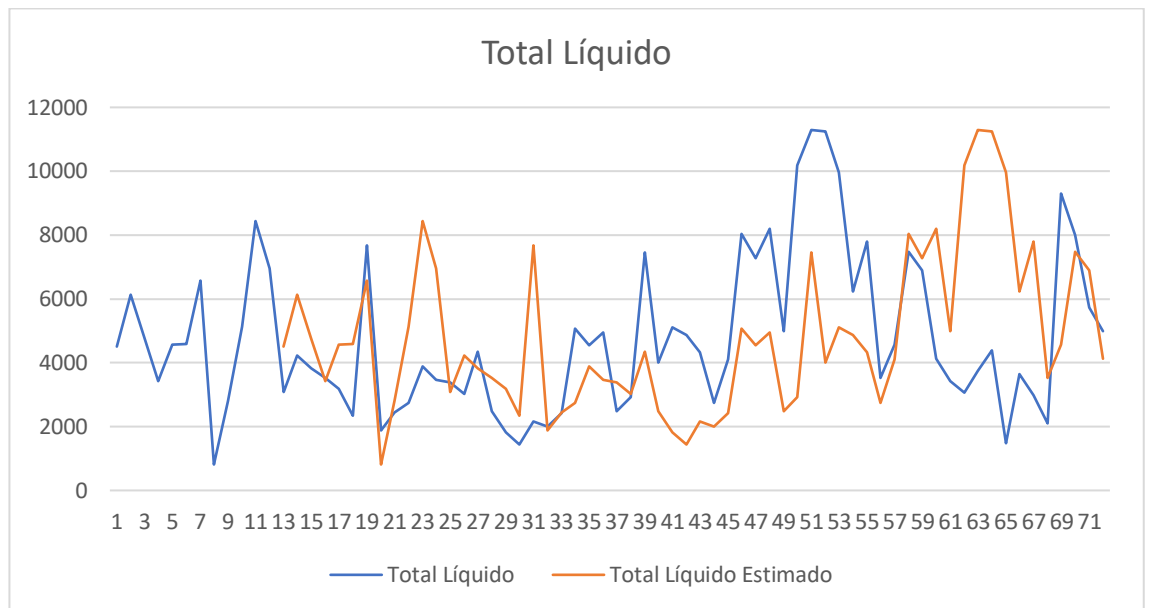
9.2.7.1. Métodos simples - Mean



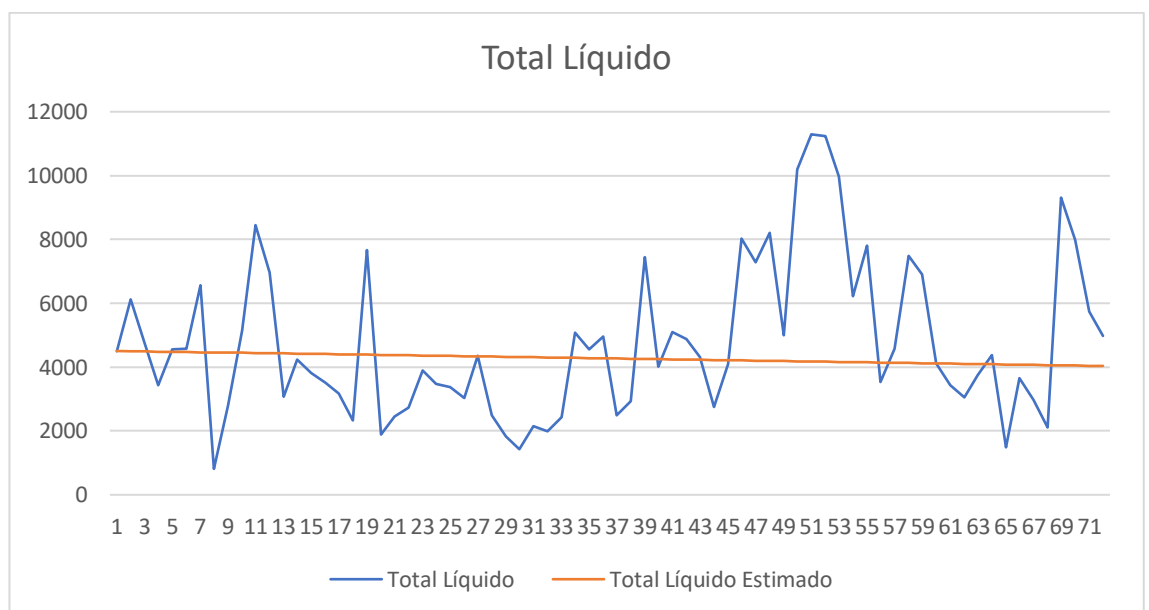
9.2.7.2. Métodos simples - Naive



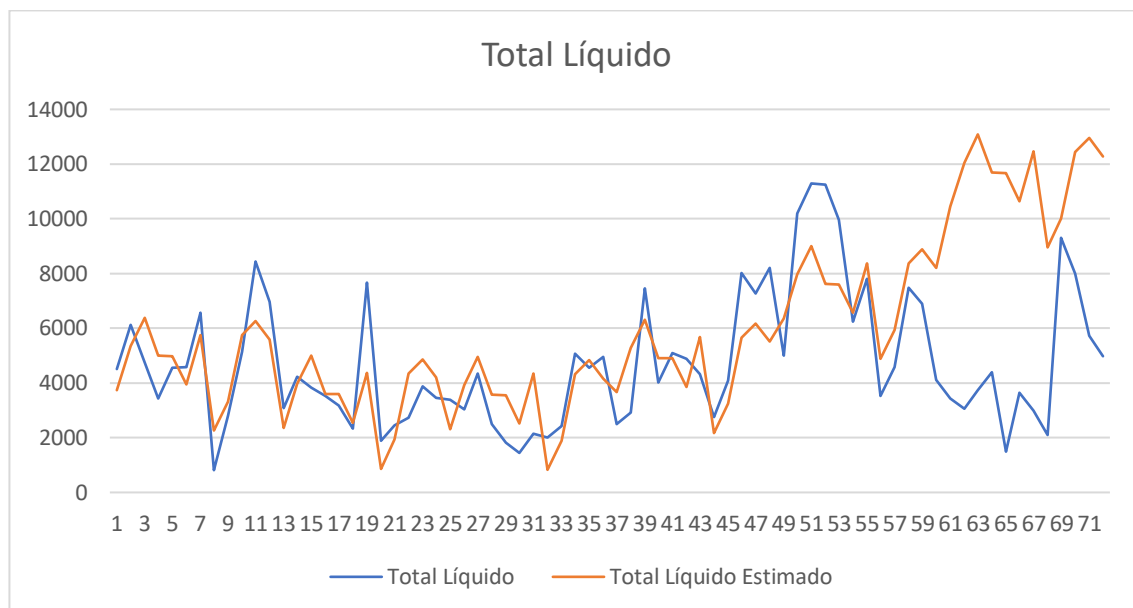
9.2.7.3. Métodos simples - SNaive



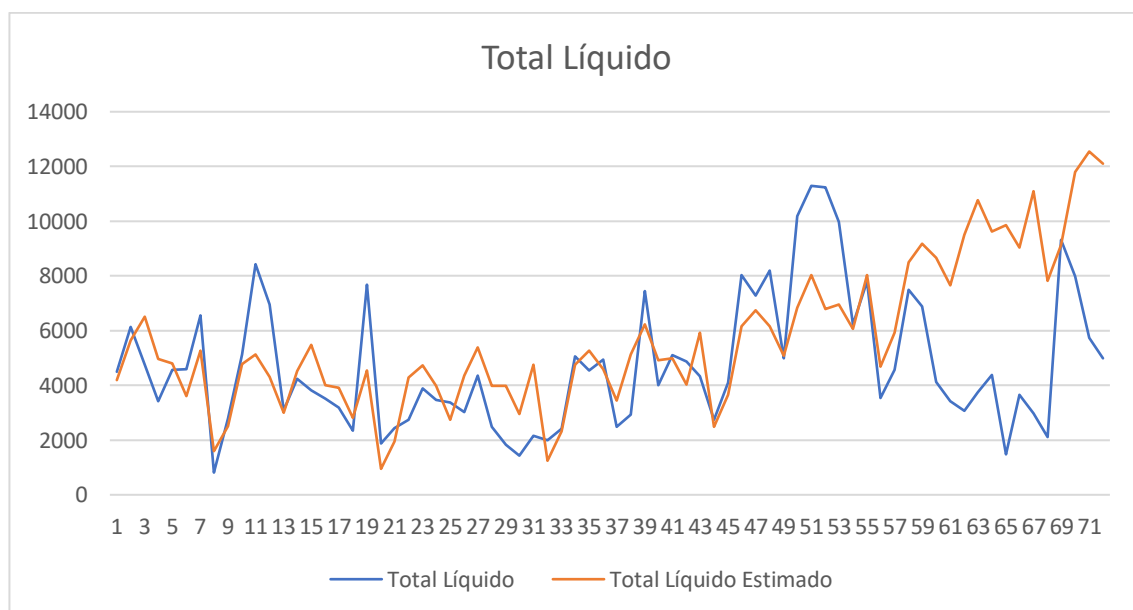
9.2.7.4. Métodos simples - Drift



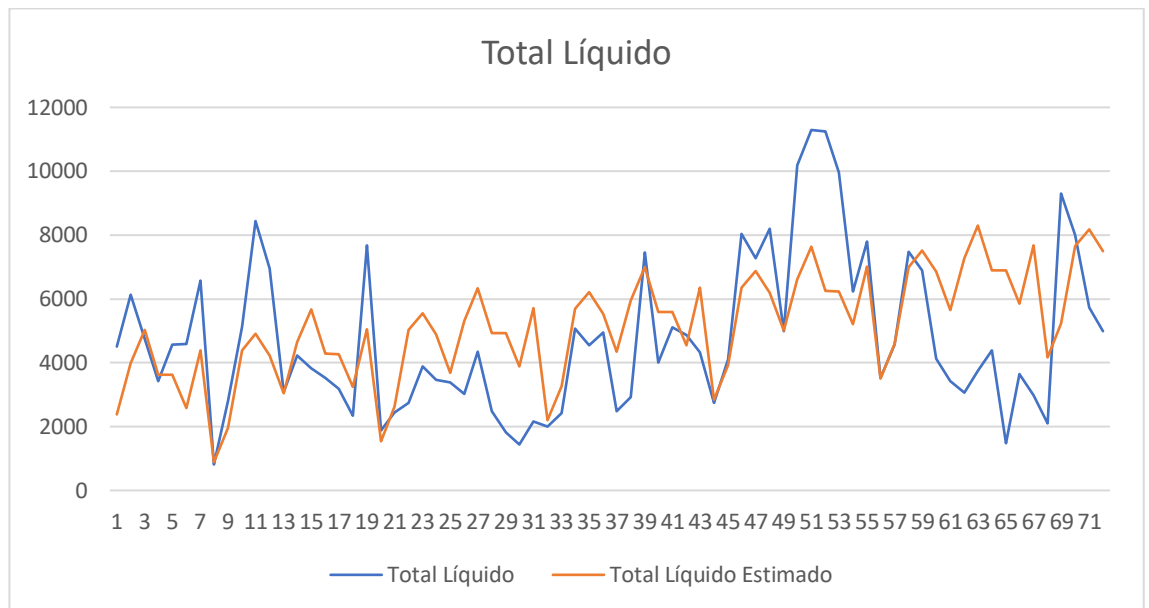
9.2.7.5. Métodos regressivos - Modelo 1



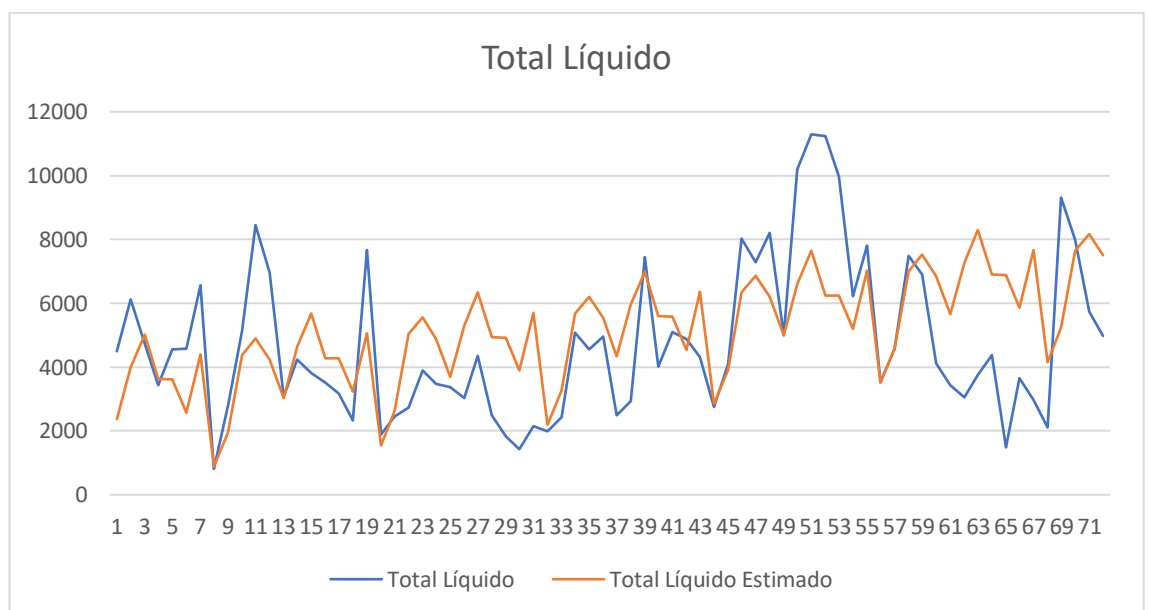
9.2.7.6. Métodos regressivos - Modelo 2



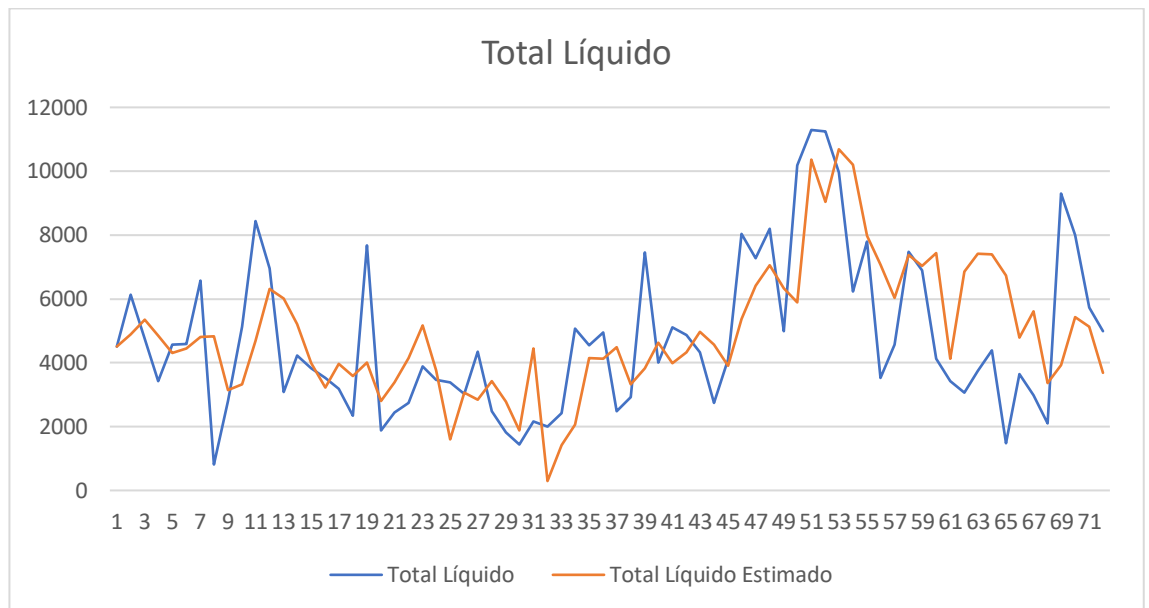
9.2.7.7. Métodos regressivos - Modelo 3



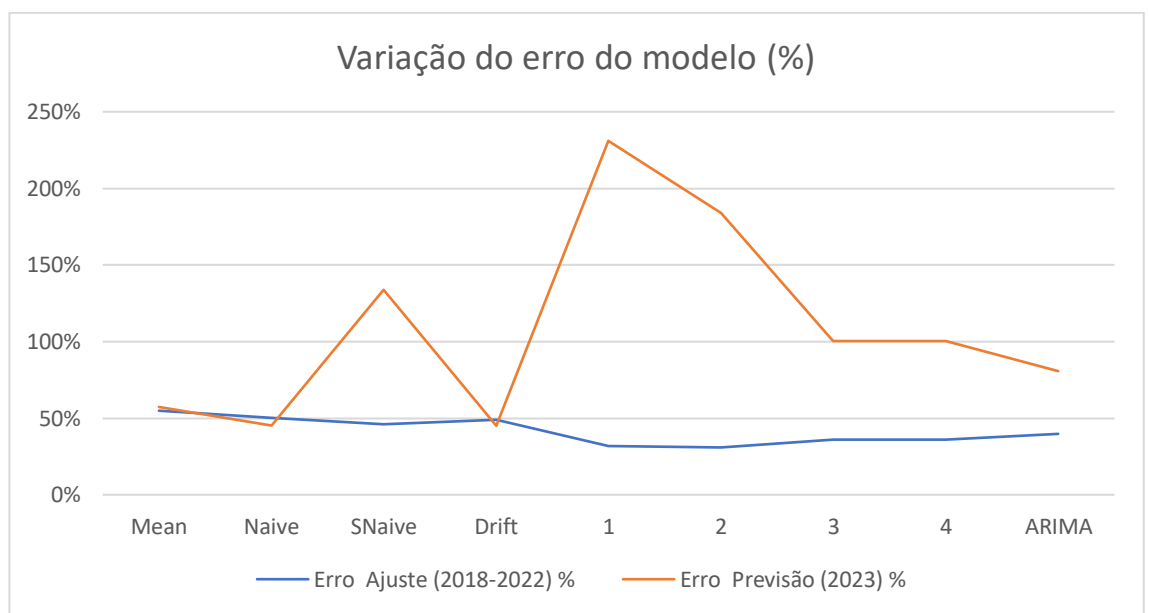
9.2.7.8. Métodos regressivos - Modelo 4



9.2.7.9. ARIMA

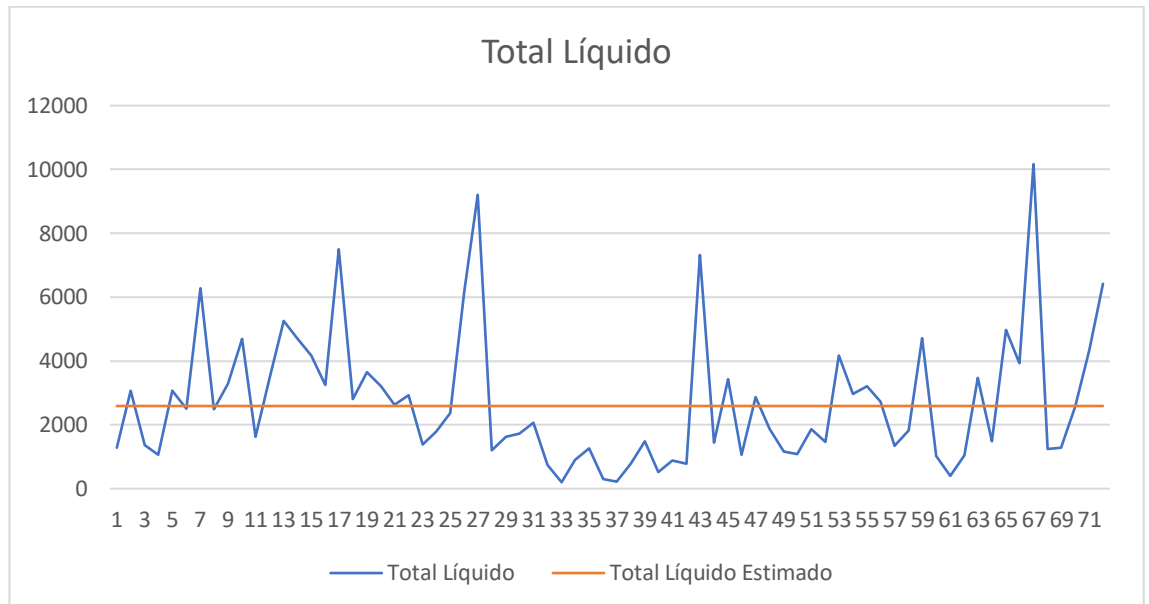


9.2.7.10. Representação gráfica dos erros dos modelos

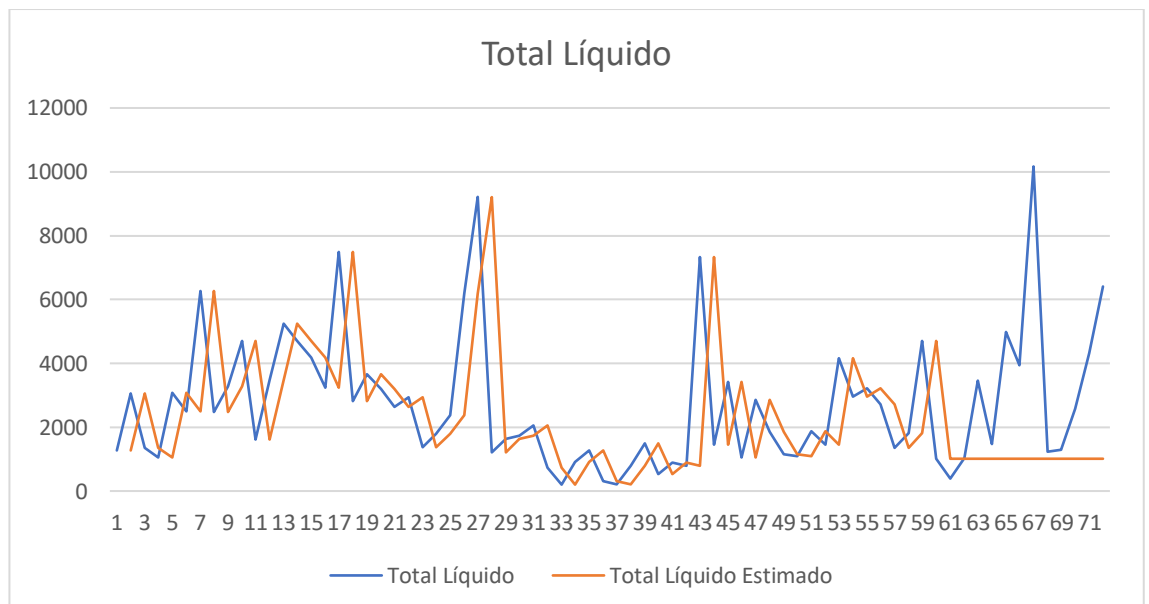


9.2.8. Total Líquido por Tipo de Lavagem – *Cluster 5*

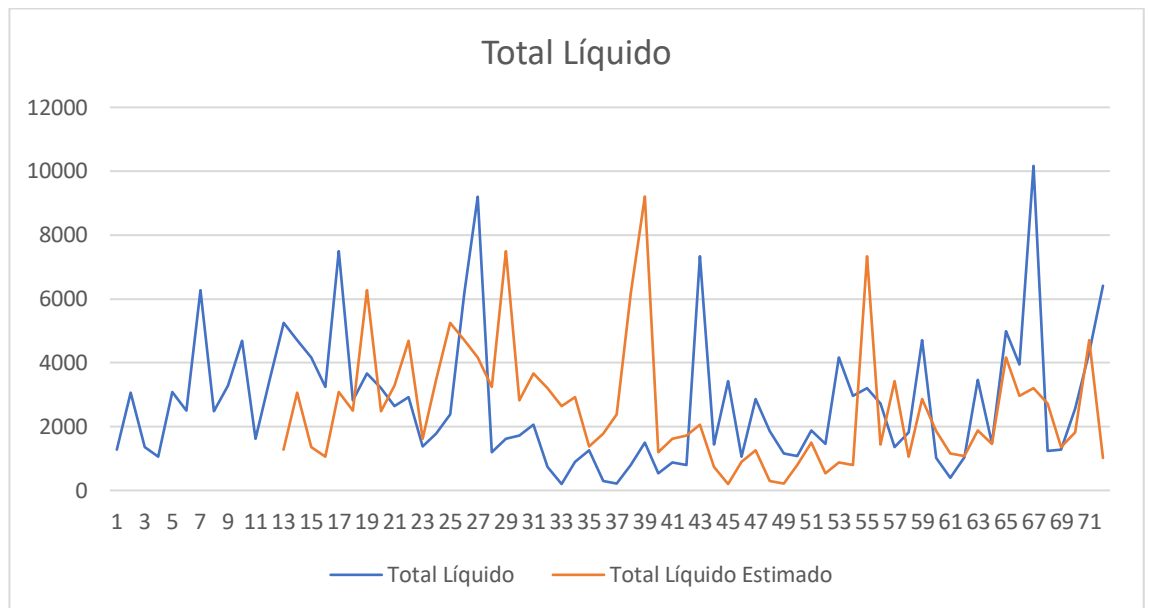
9.2.8.1. Métodos simples - Mean



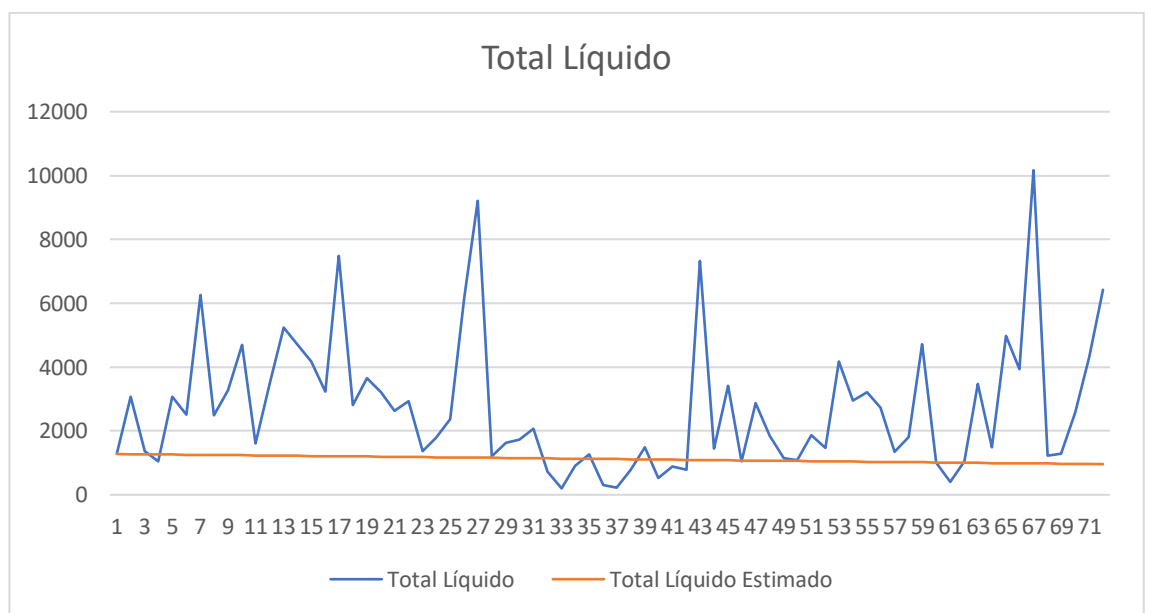
9.2.8.2. Métodos simples - Naive



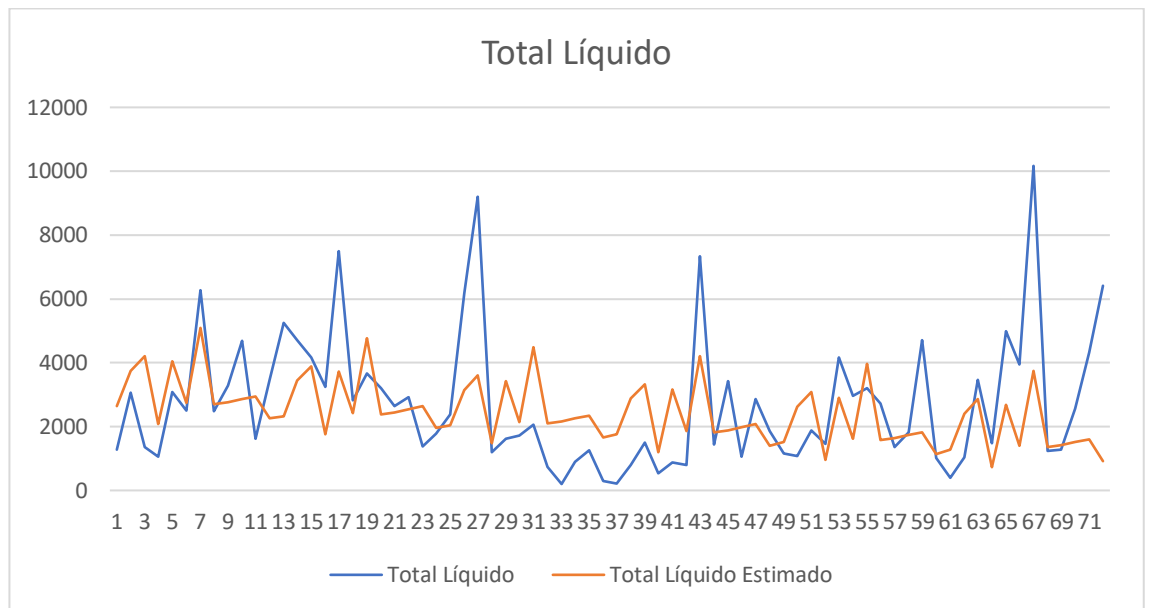
9.2.8.3. Métodos simples - SNaive



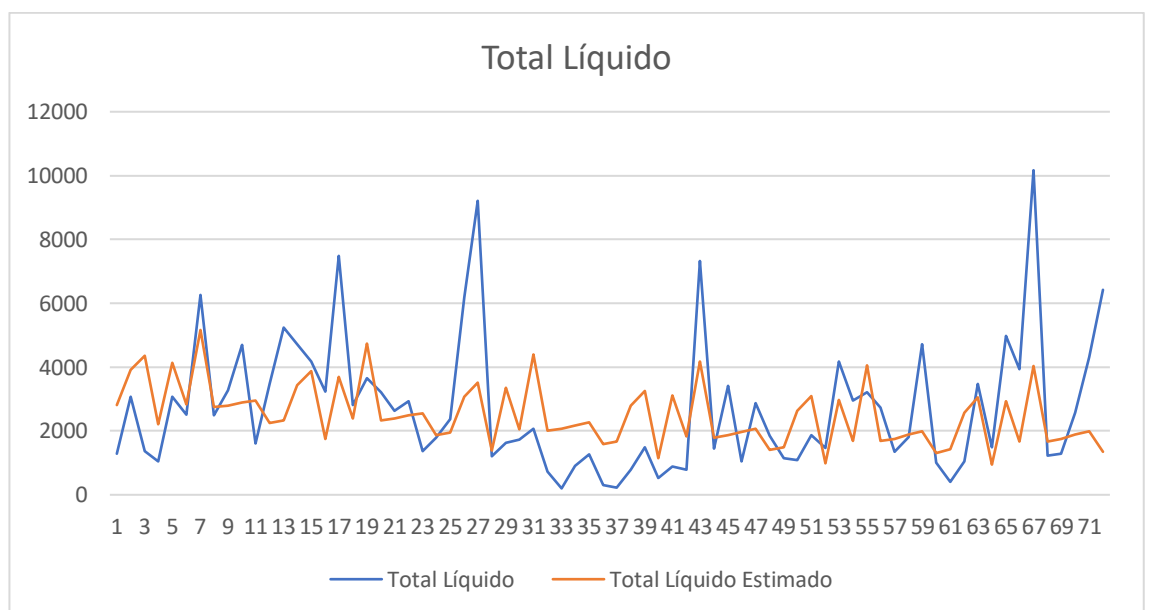
9.2.8.4. Métodos simples - Drift



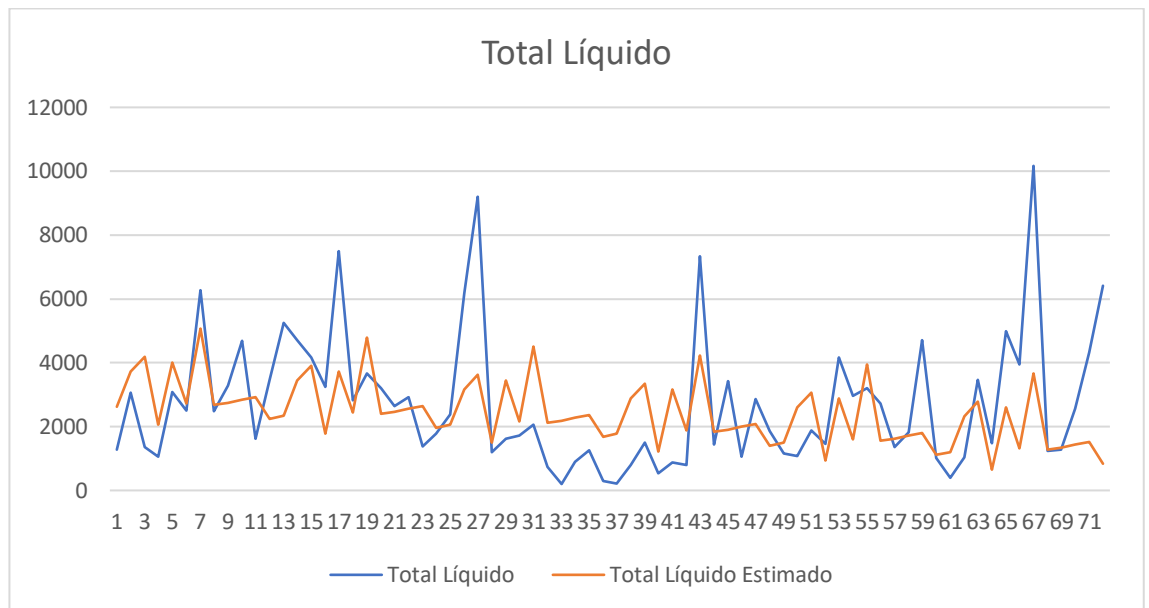
9.2.8.5. Métodos regressivos - Modelo 1



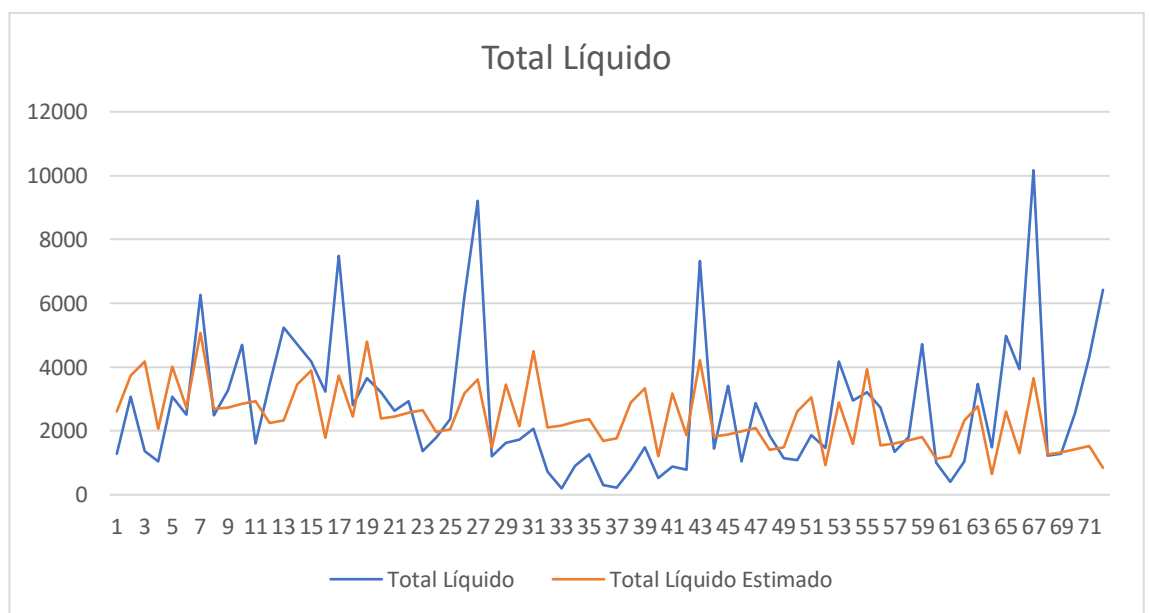
9.2.8.6. Métodos regressivos - Modelo 2



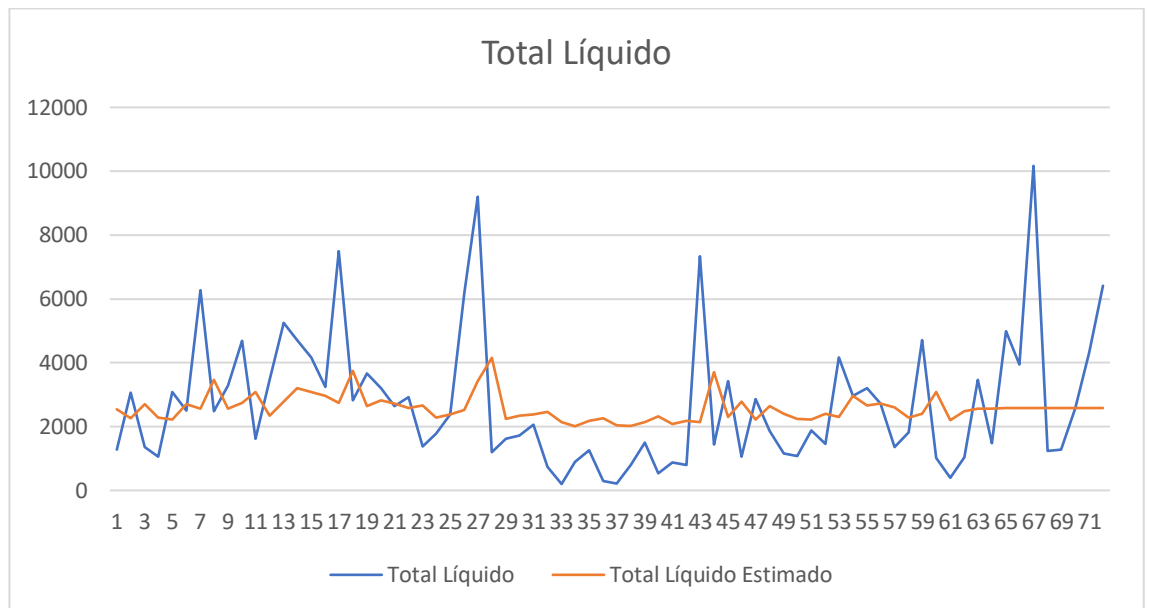
9.2.8.7. Métodos regressivos - Modelo 3



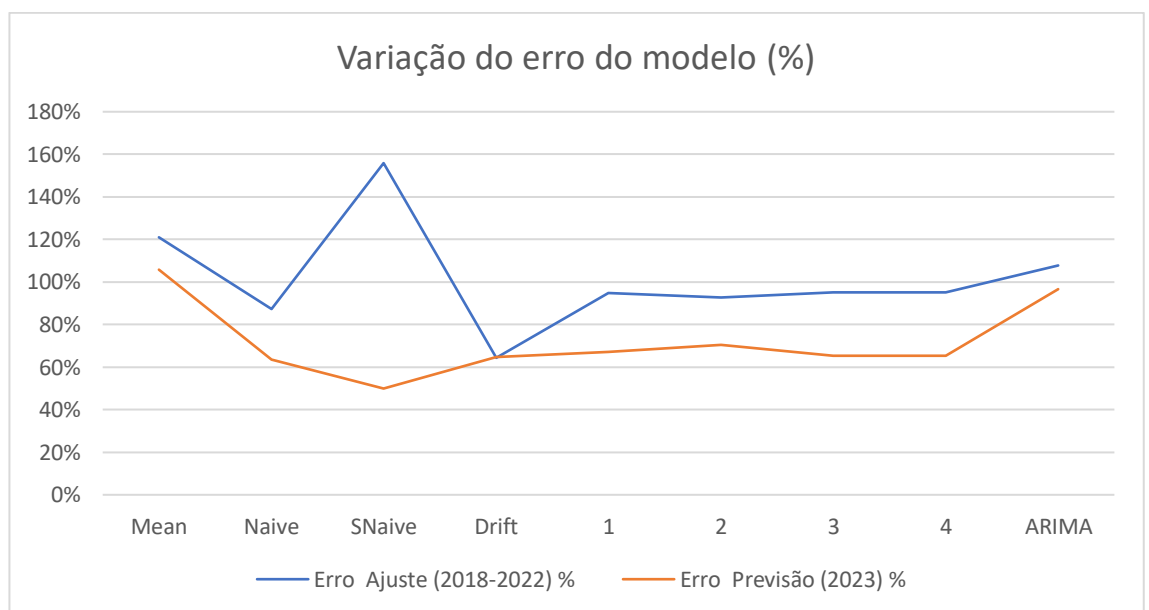
9.2.8.8. Métodos regressivos - Modelo 4



9.2.8.9. ARIMA

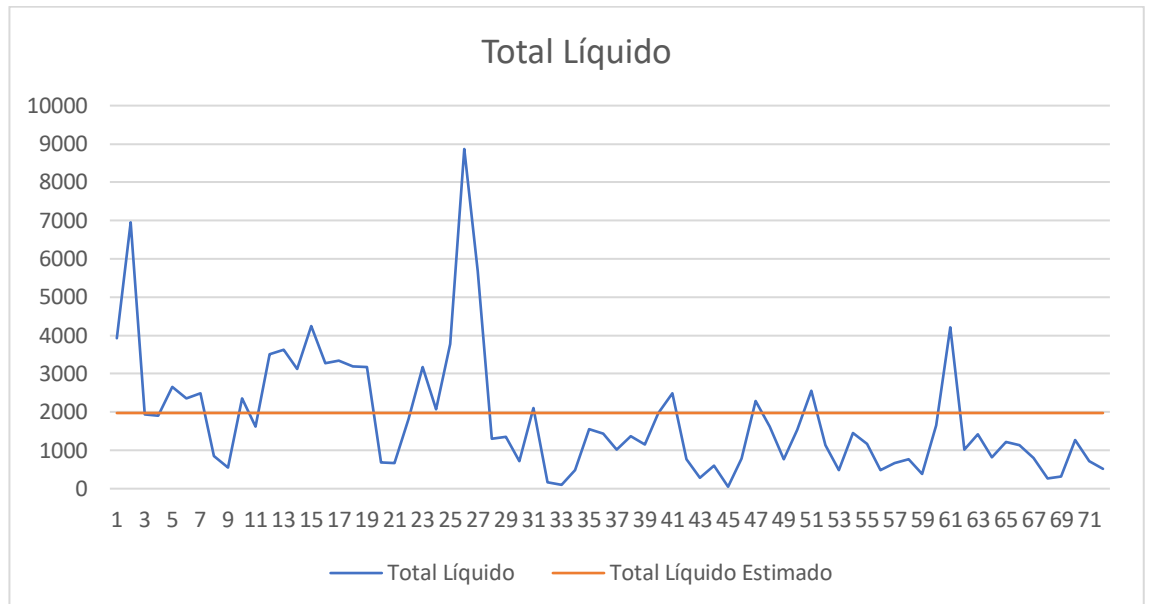


9.2.8.10. Representação gráfica dos erros dos modelos

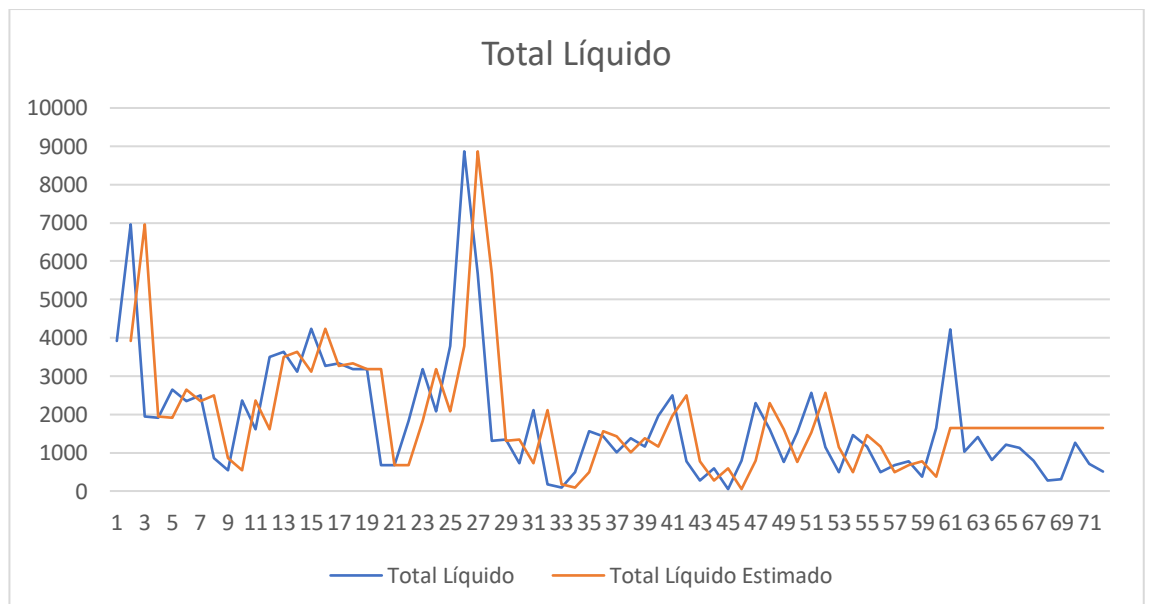


9.2.9. Total Líquido por Tipo de Lavagem – *Cluster 6*

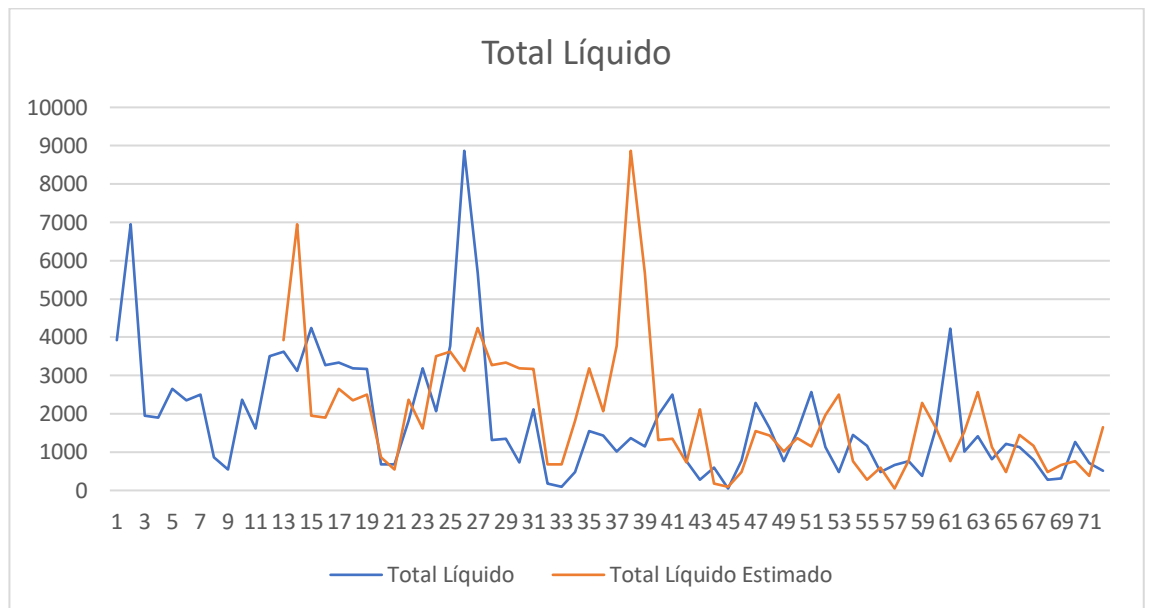
9.2.9.1. Métodos simples - Mean



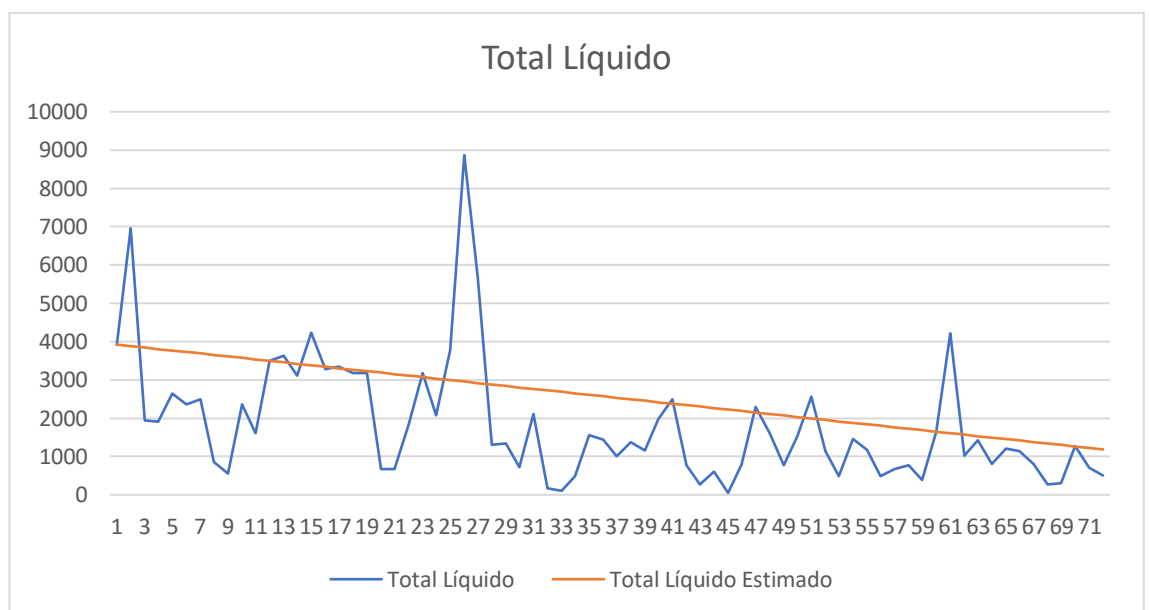
9.2.9.2. Métodos simples - Naive



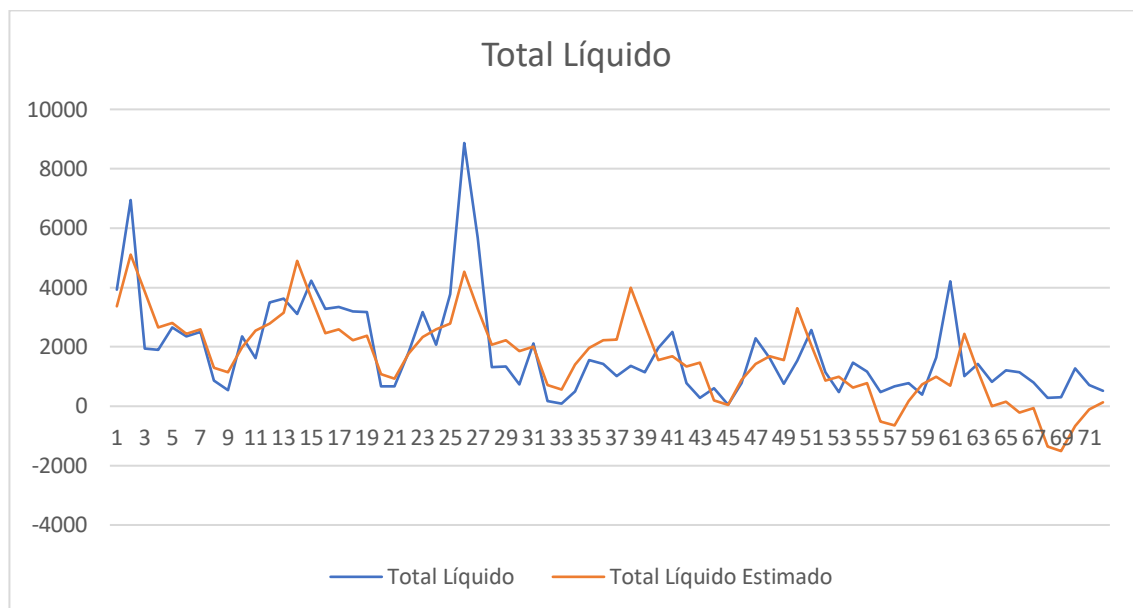
9.2.9.3. Métodos simples - SNaive



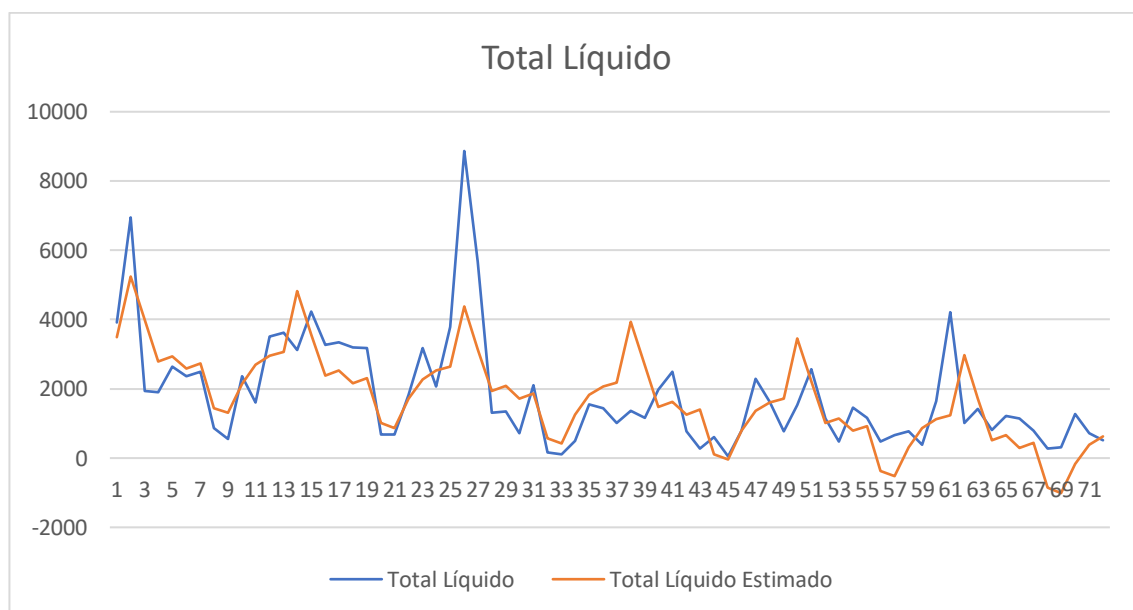
9.2.9.4. Métodos simples - Drift



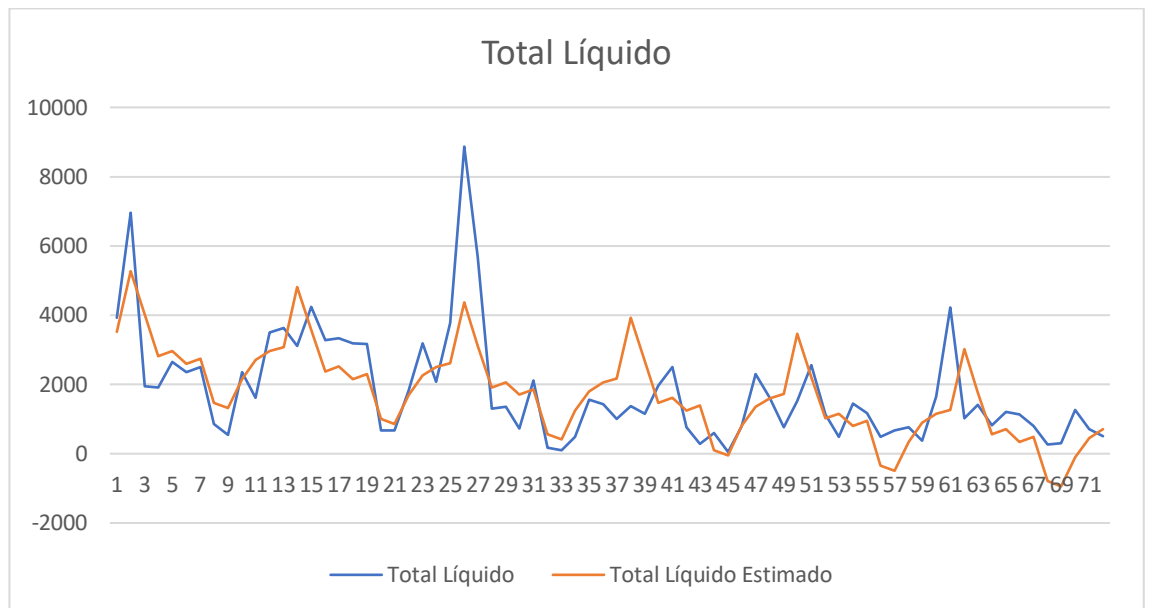
9.2.9.5. Métodos regressivos - Modelo 1



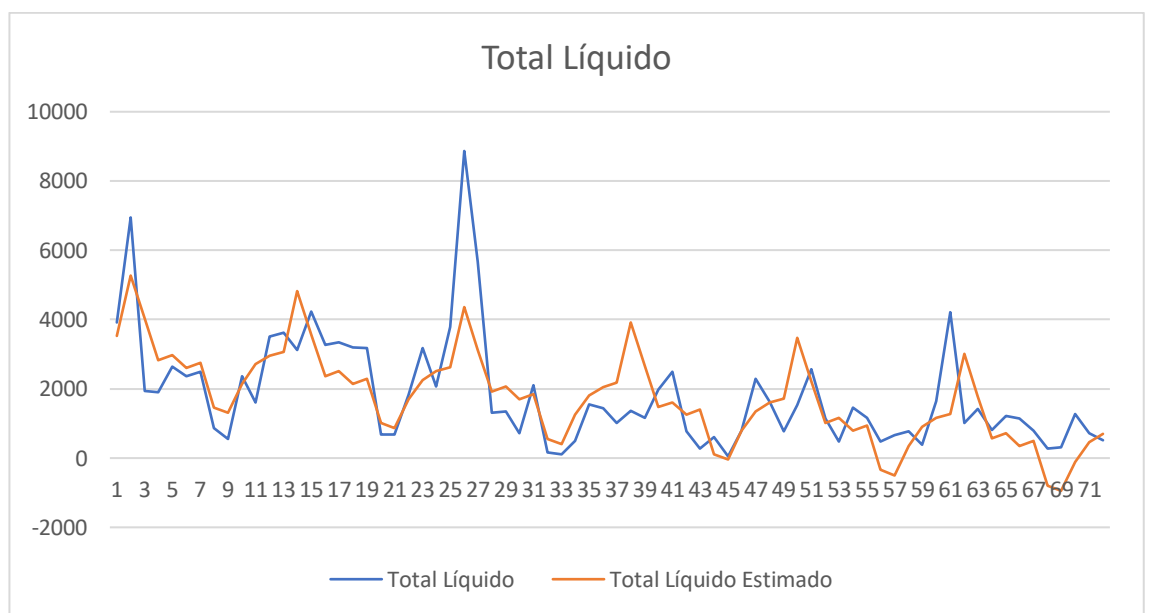
9.2.9.6. Métodos regressivos - Modelo 2



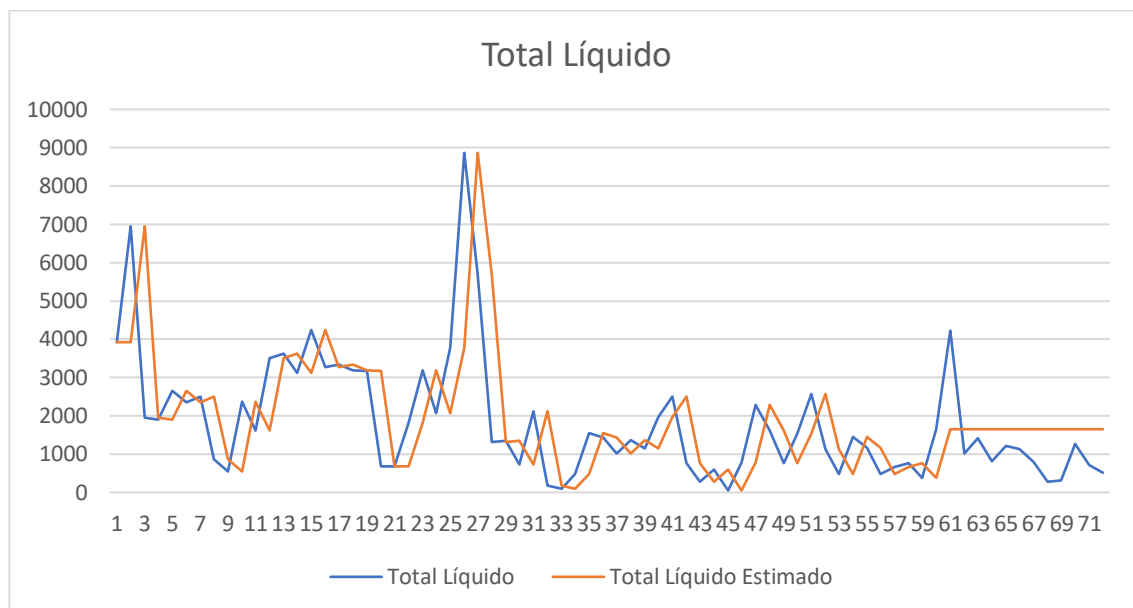
9.2.9.7. Métodos regressivos - Modelo 3



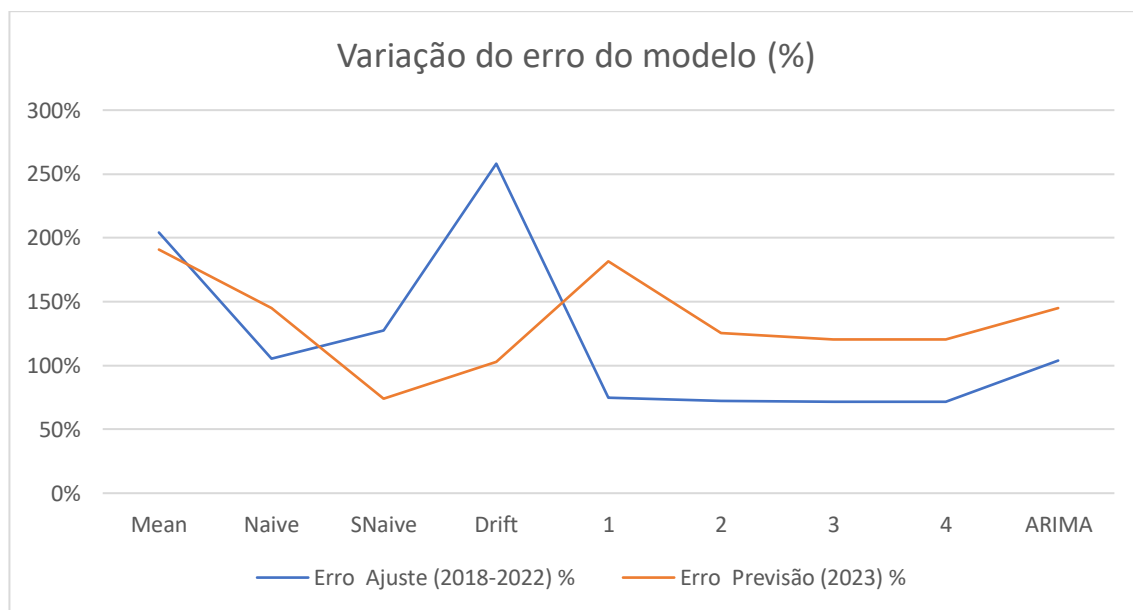
9.2.9.8. Métodos regressivos - Modelo 4



9.2.9.9. ARIMA

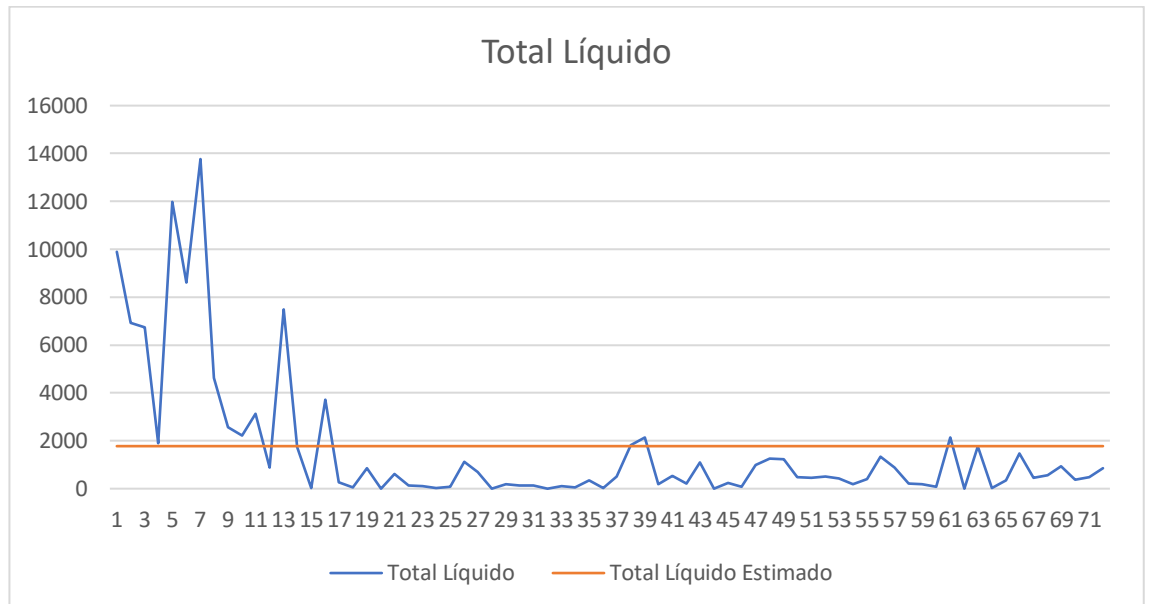


9.2.9.10. Representação gráfica dos erros dos modelos

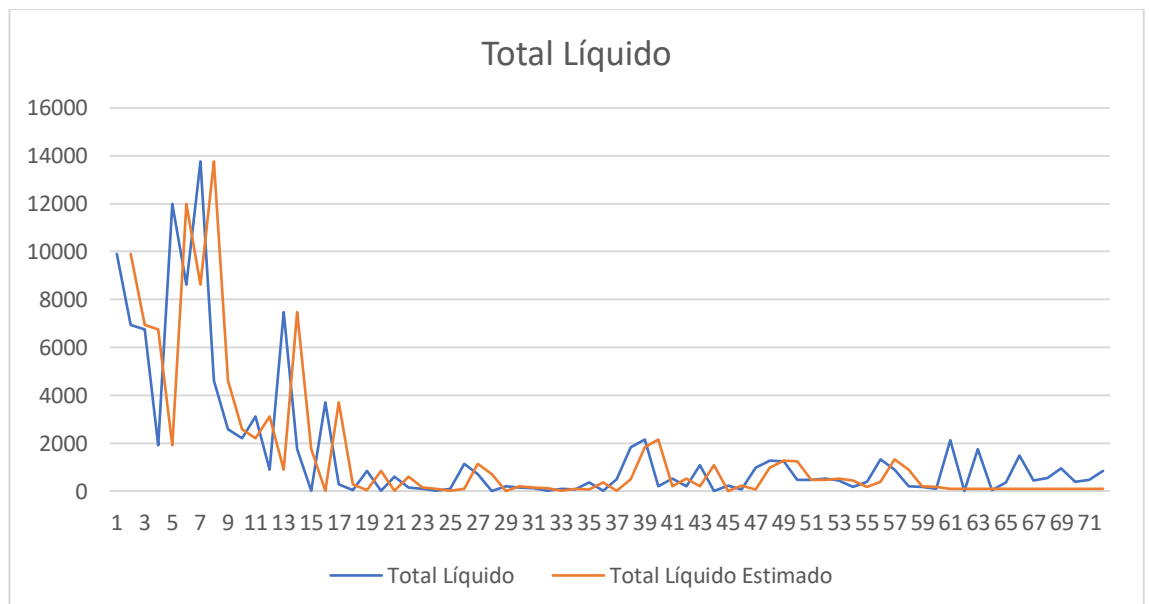


9.2.10. Total Líquido por Tipo de Lavagem – *Cluster 7*

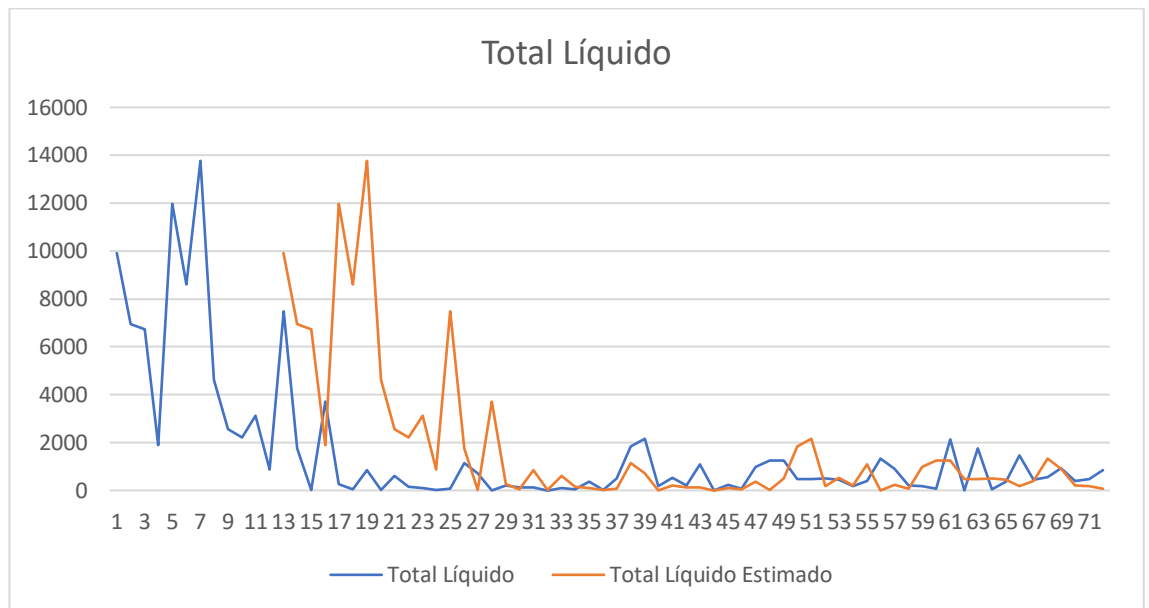
9.2.10.1. Métodos simples - Mean



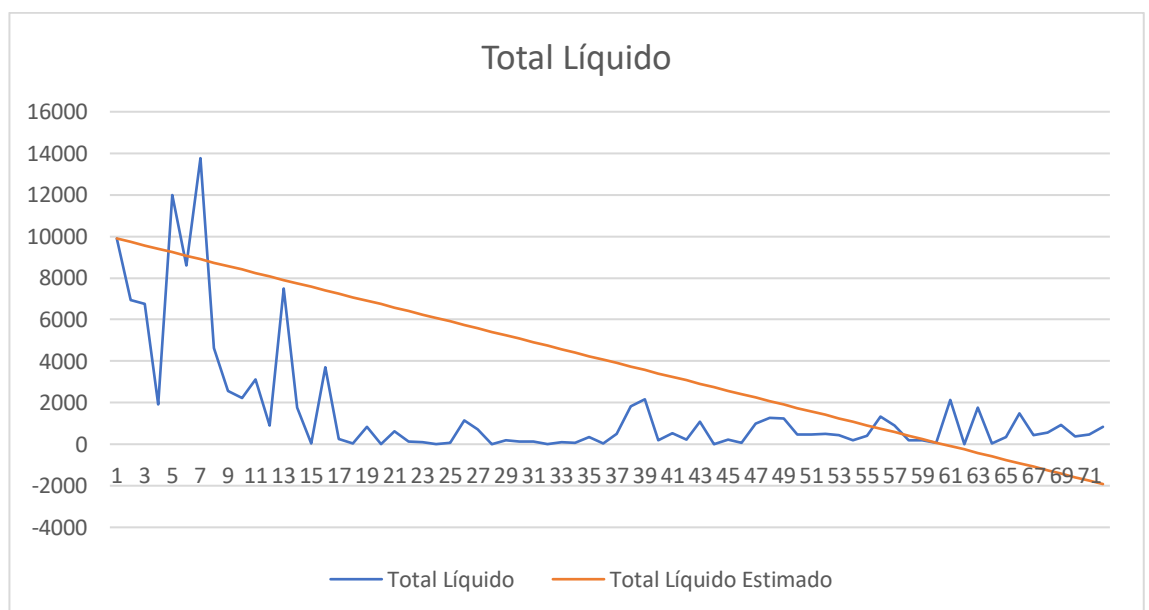
9.2.10.2. Métodos simples - Naive



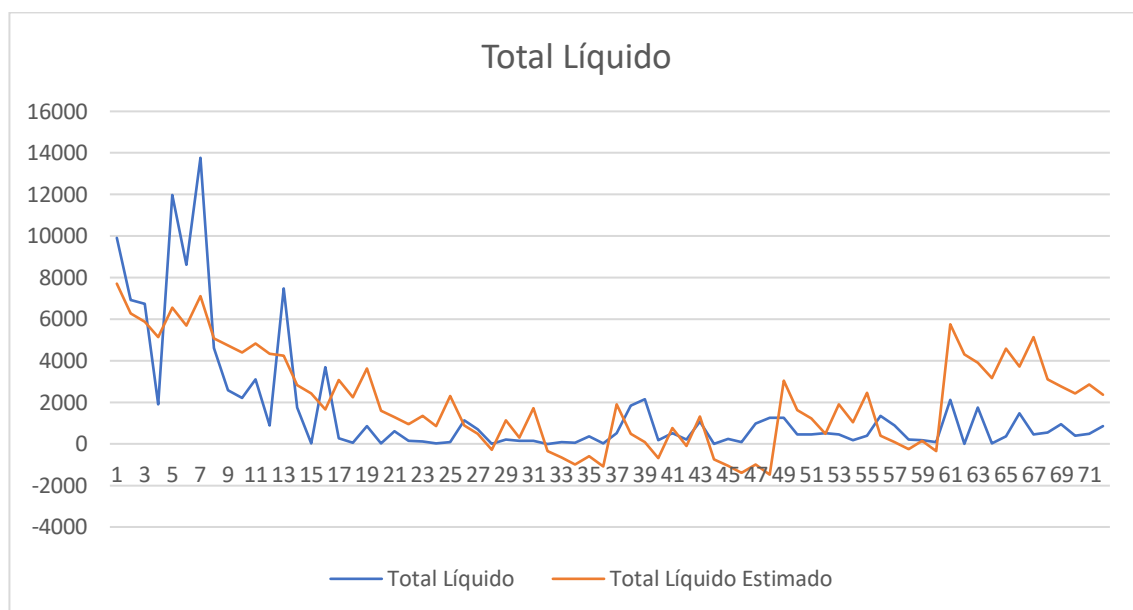
9.2.10.3. Métodos simples - SNaive



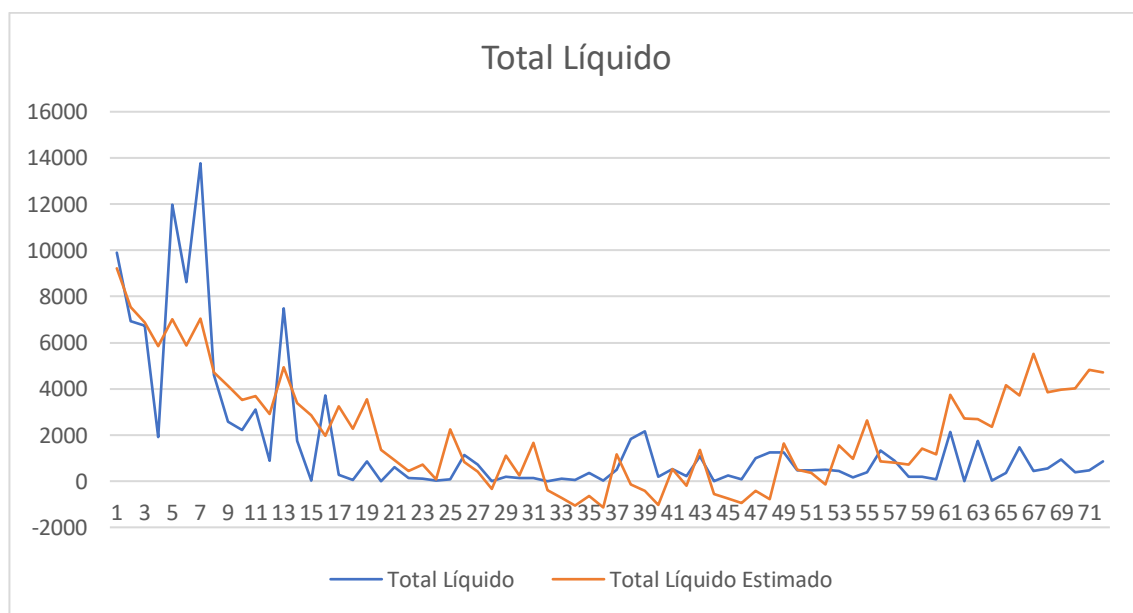
9.2.10.4. Métodos simples - Drift



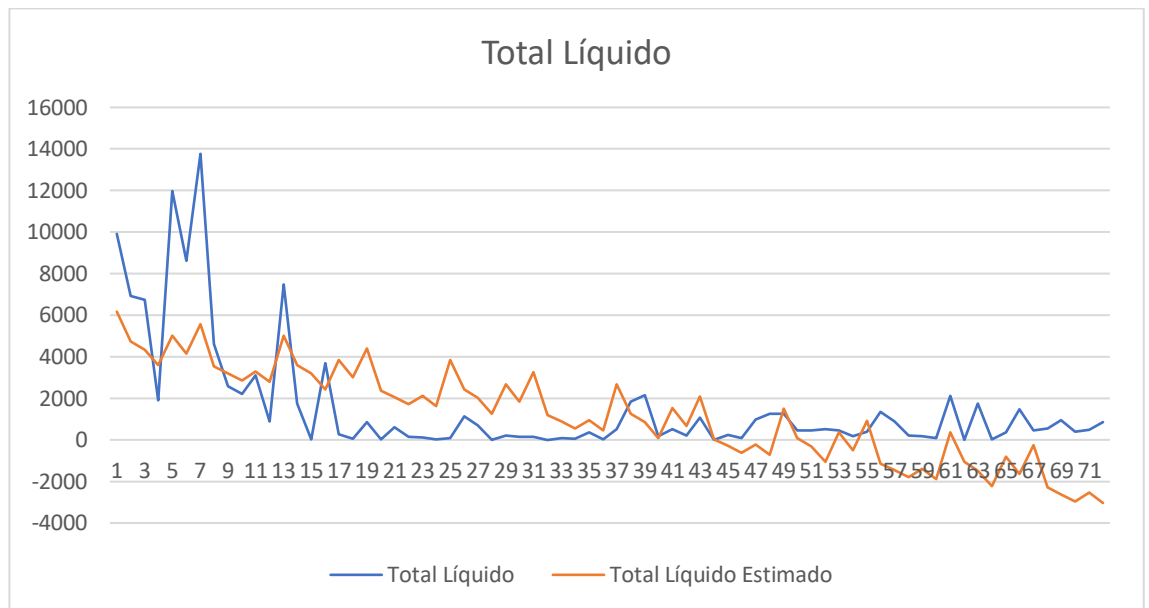
9.2.10.5. Métodos regressivos - Modelo 1



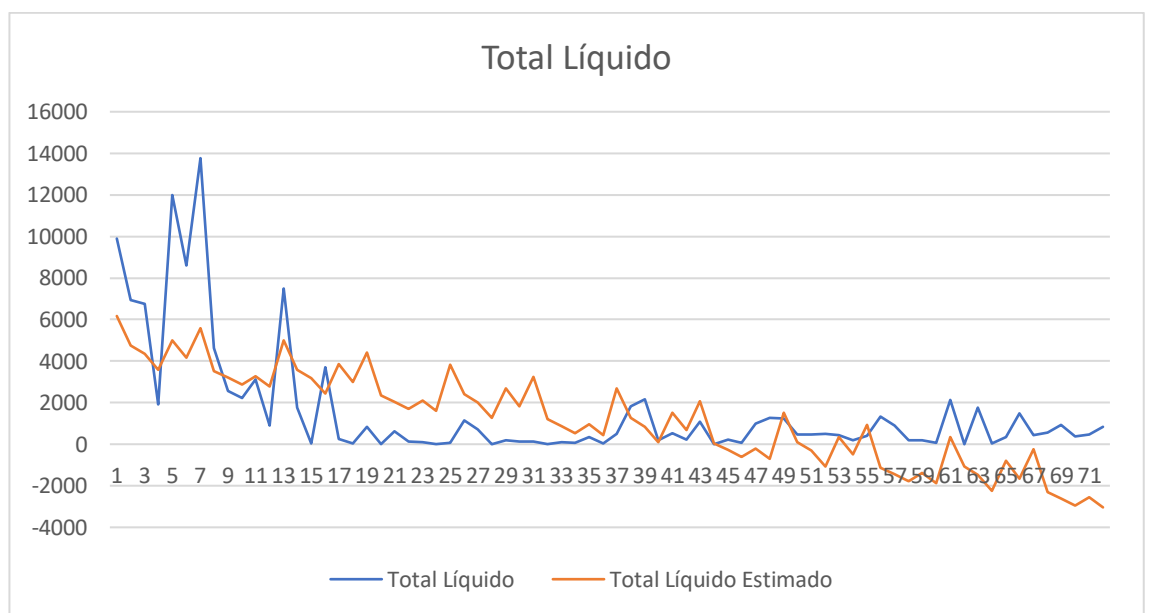
9.2.10.6. Métodos regressivos - Modelo 2



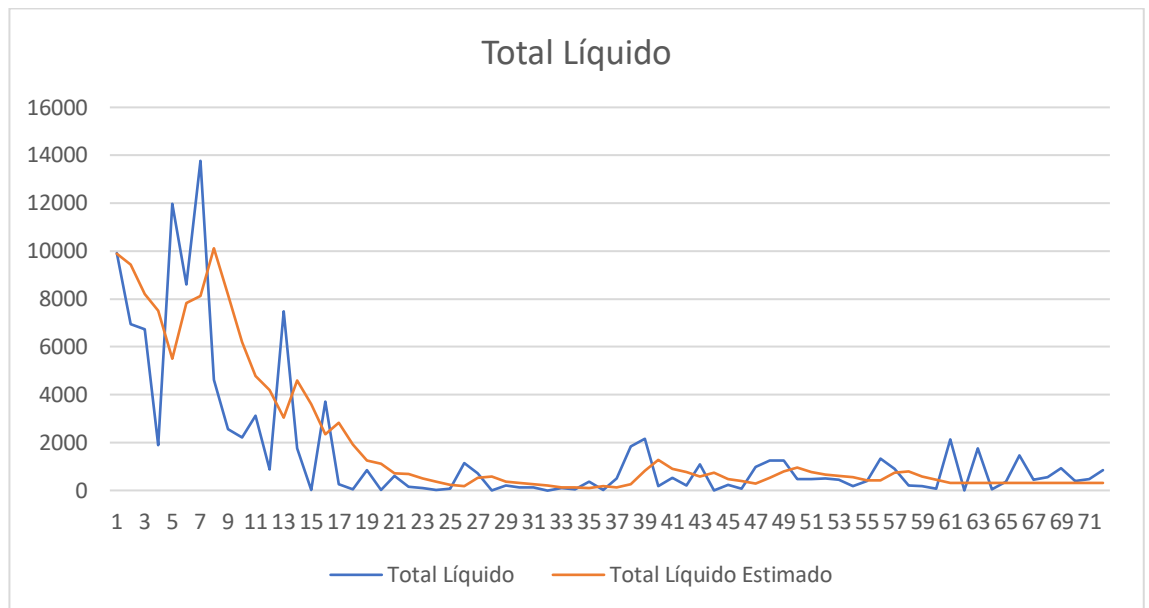
9.2.10.7. Métodos regressivos - Modelo 3



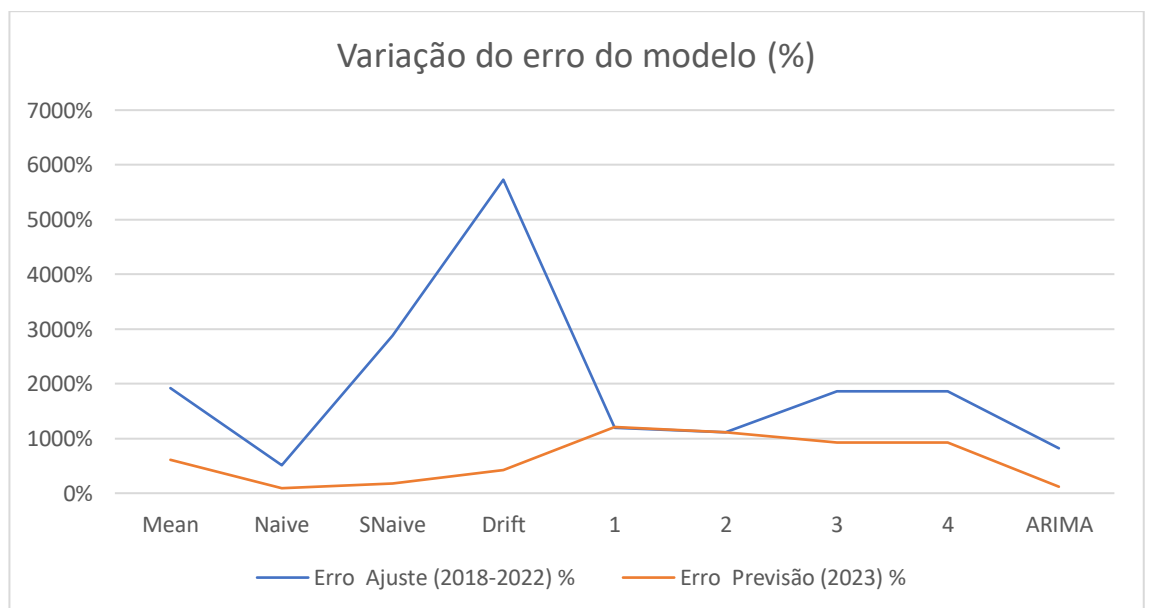
9.2.10.8. Métodos regressivos - Modelo 4



9.2.10.9. ARIMA

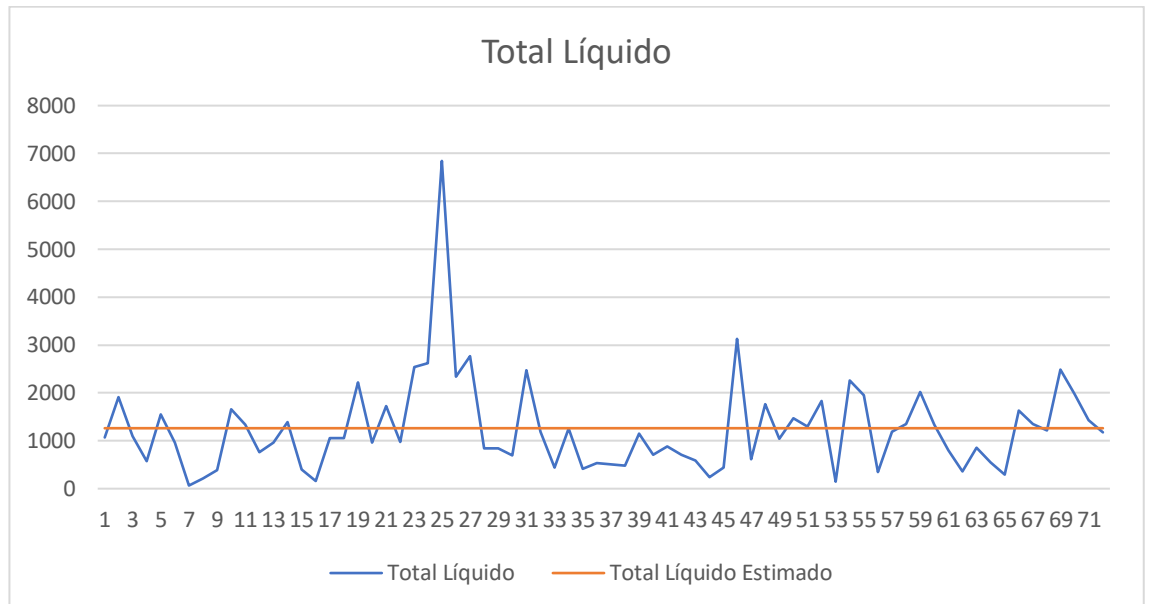


9.2.10.10. Representação gráfica dos erros dos modelos

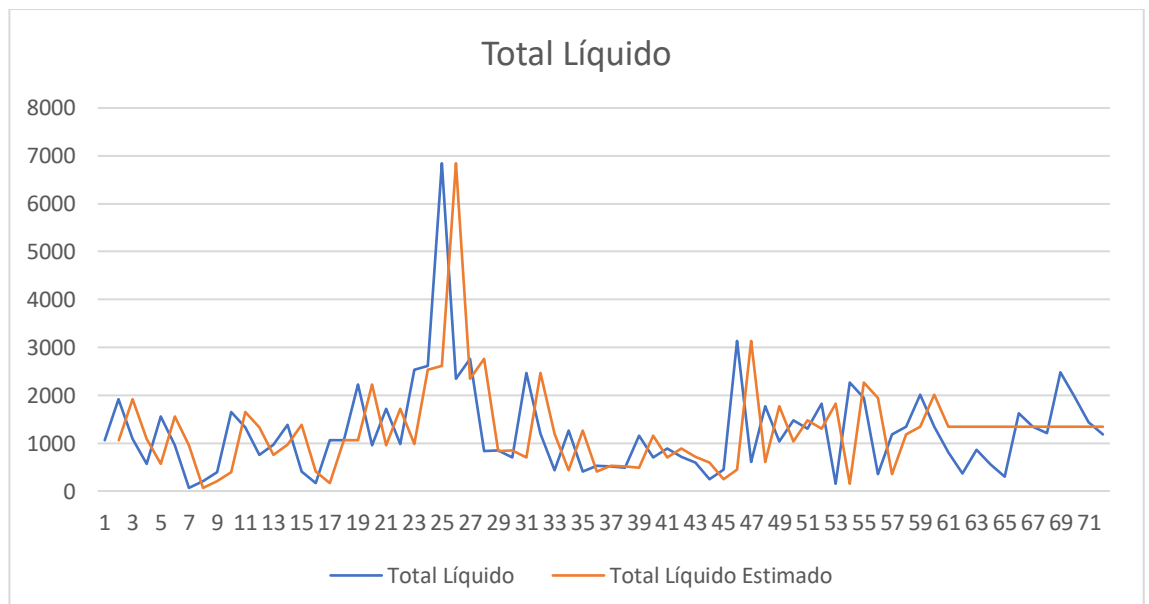


9.2.11. Total Líquido por Tipo de Lavagem – *Cluster 8*

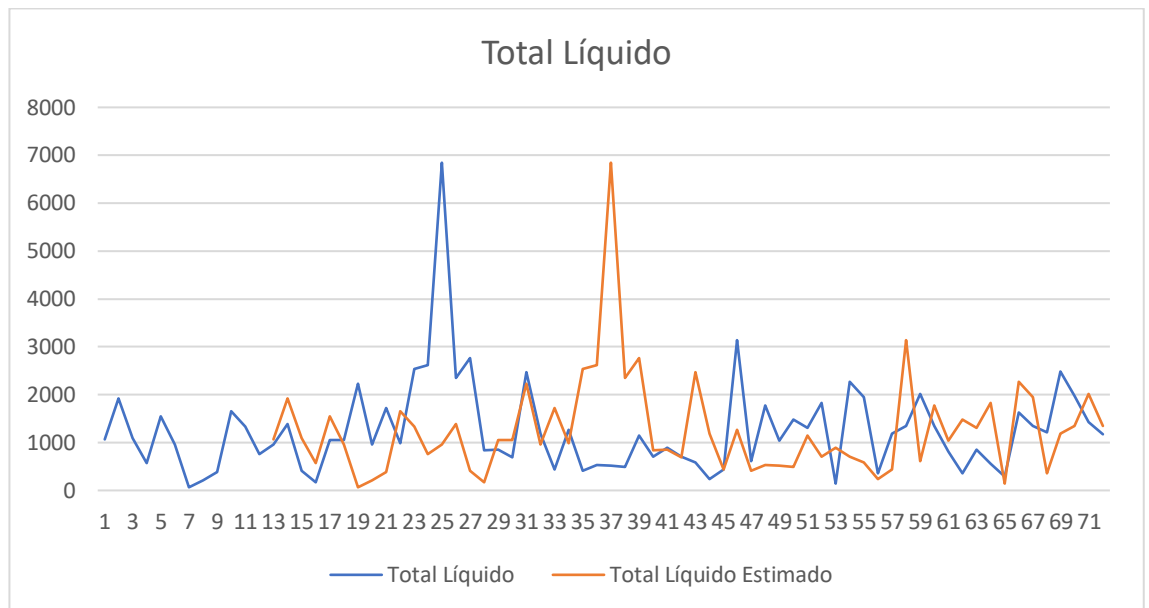
9.2.11.1. Métodos simples - Mean



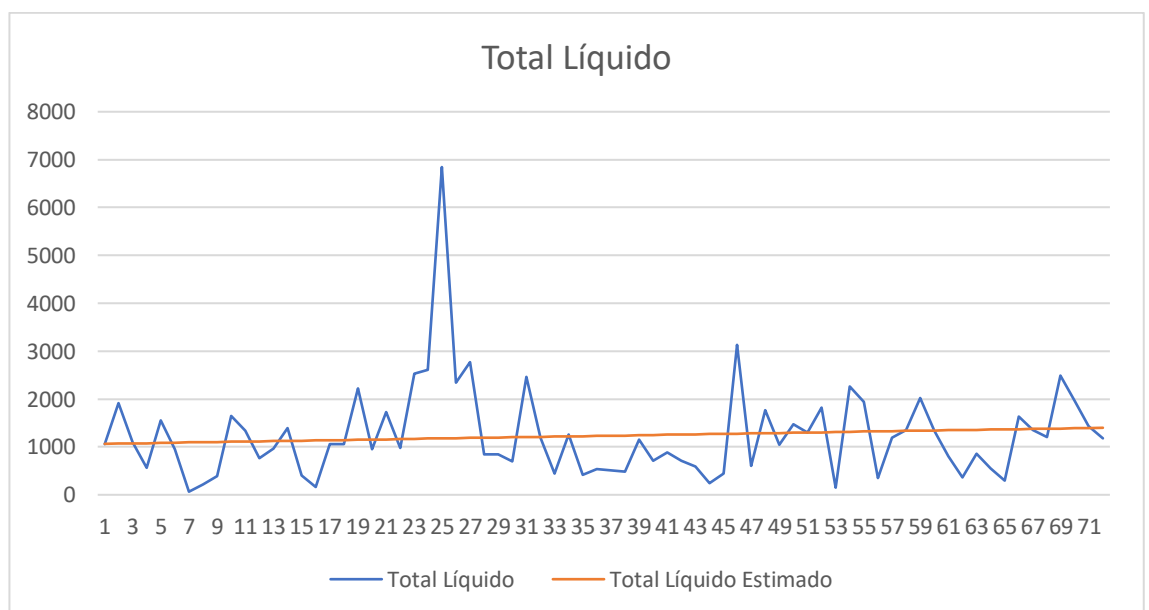
9.2.11.2. Métodos simples - Naive



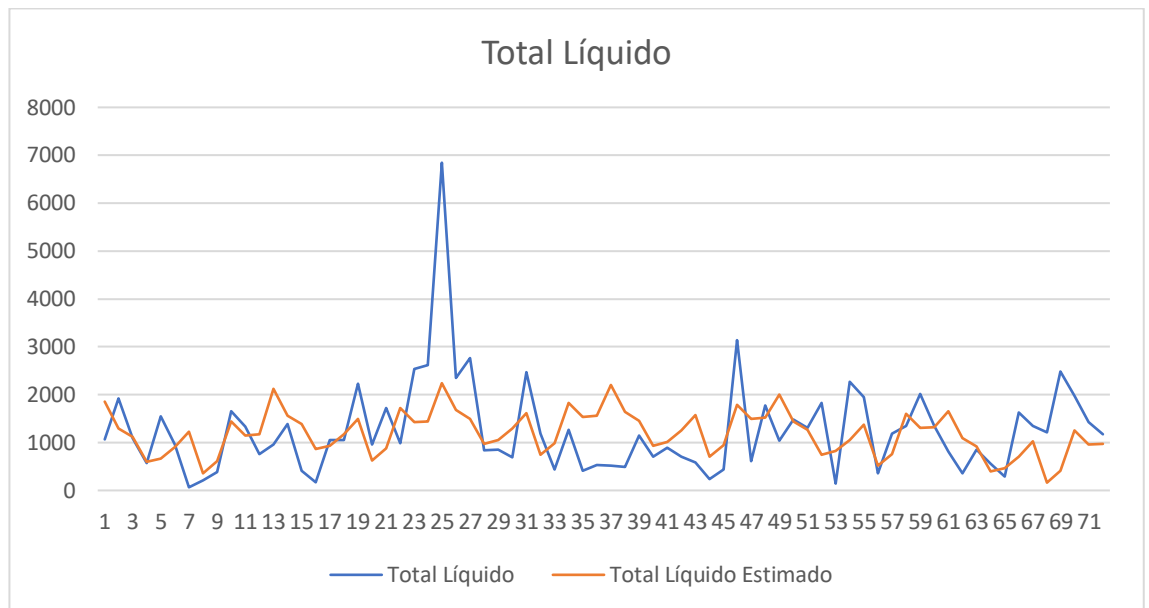
9.2.11.3. Métodos simples - SNaive



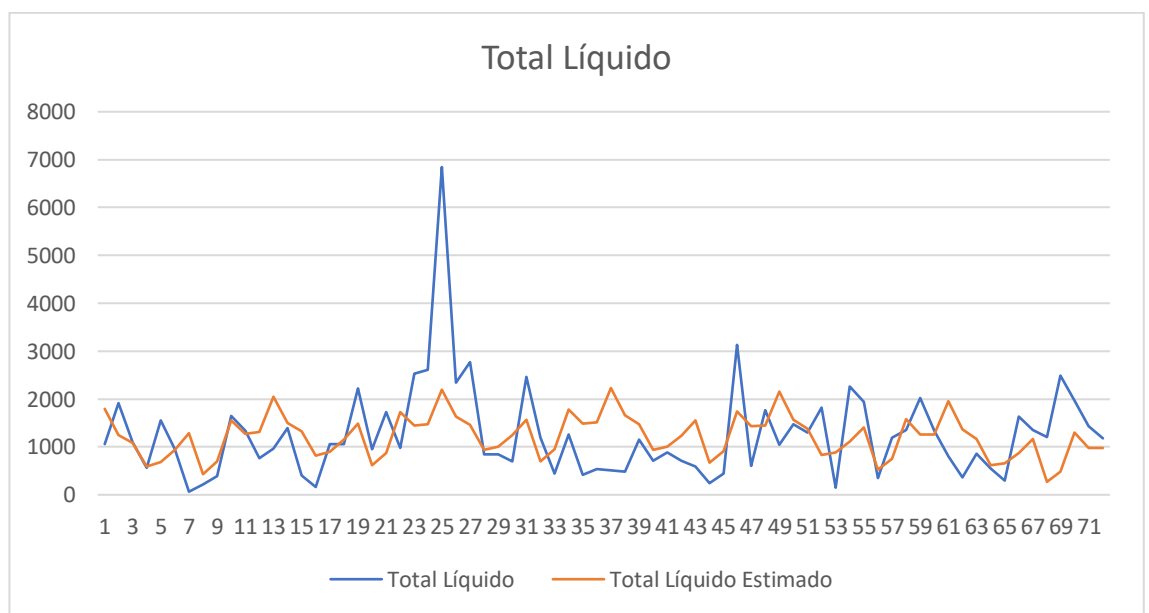
9.2.11.4. Métodos simples - Drift



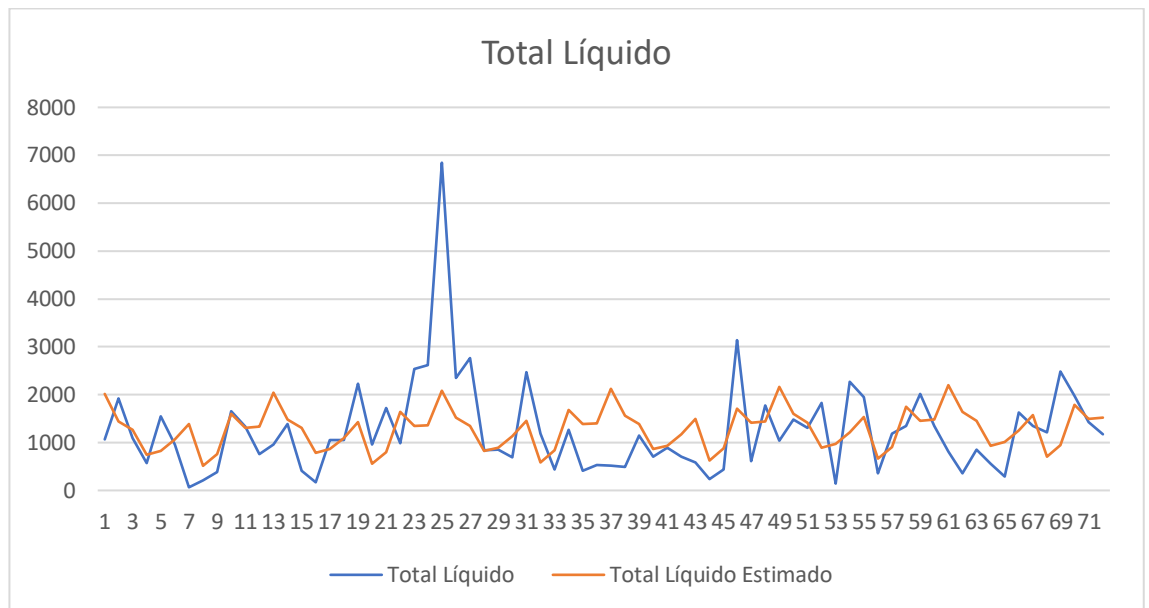
9.2.11.5. Métodos regressivos - Modelo 1



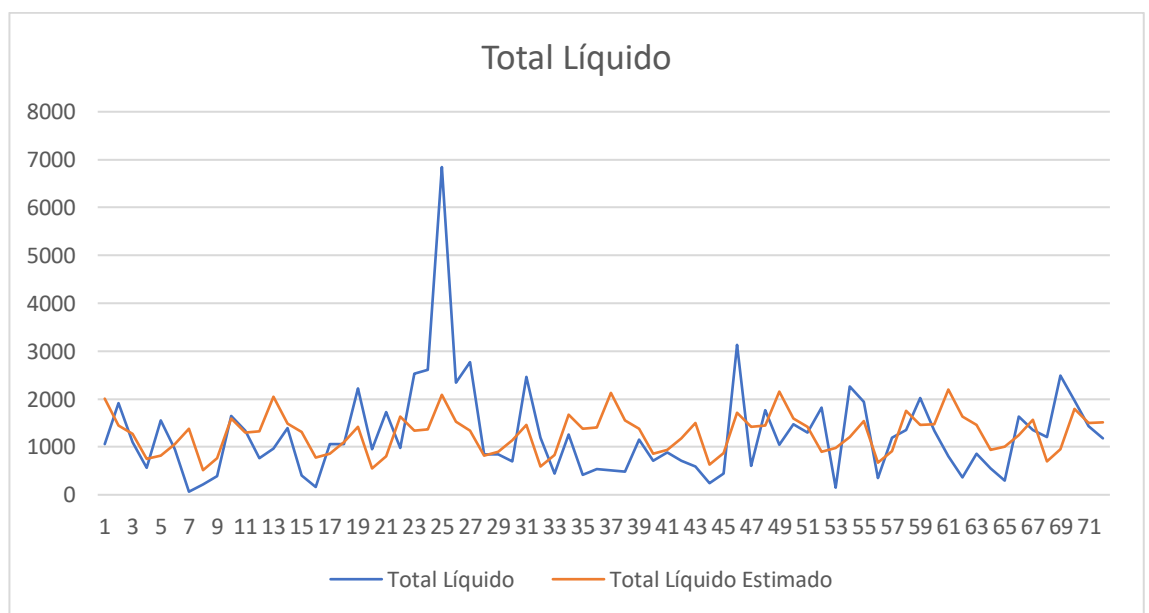
9.2.11.6. Métodos regressivos - Modelo 2



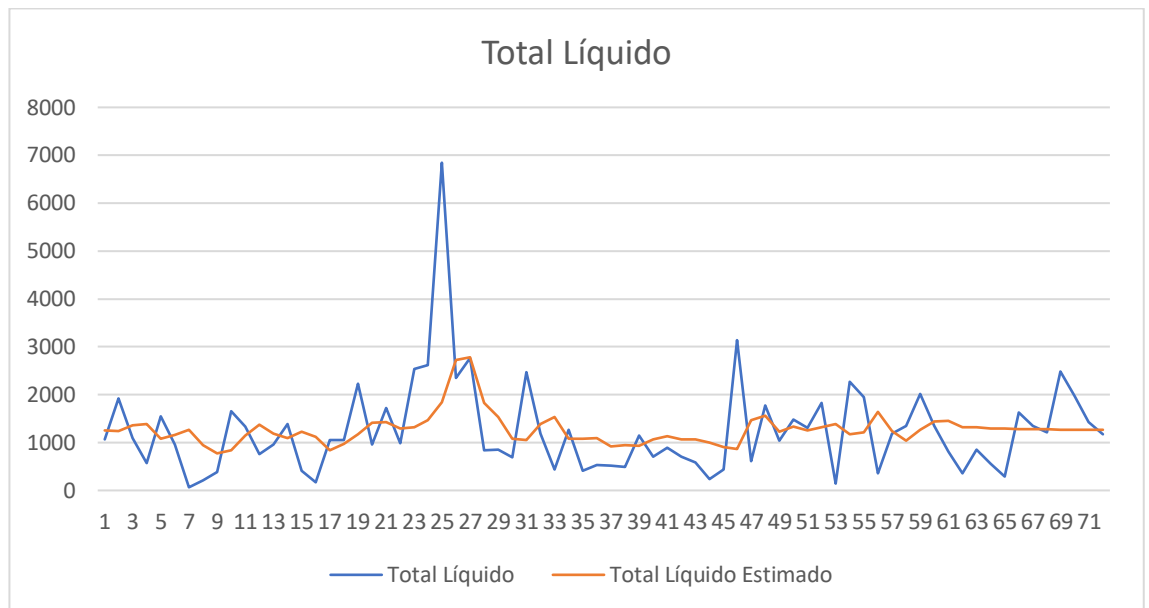
9.2.11.7. Métodos regressivos - Modelo 3



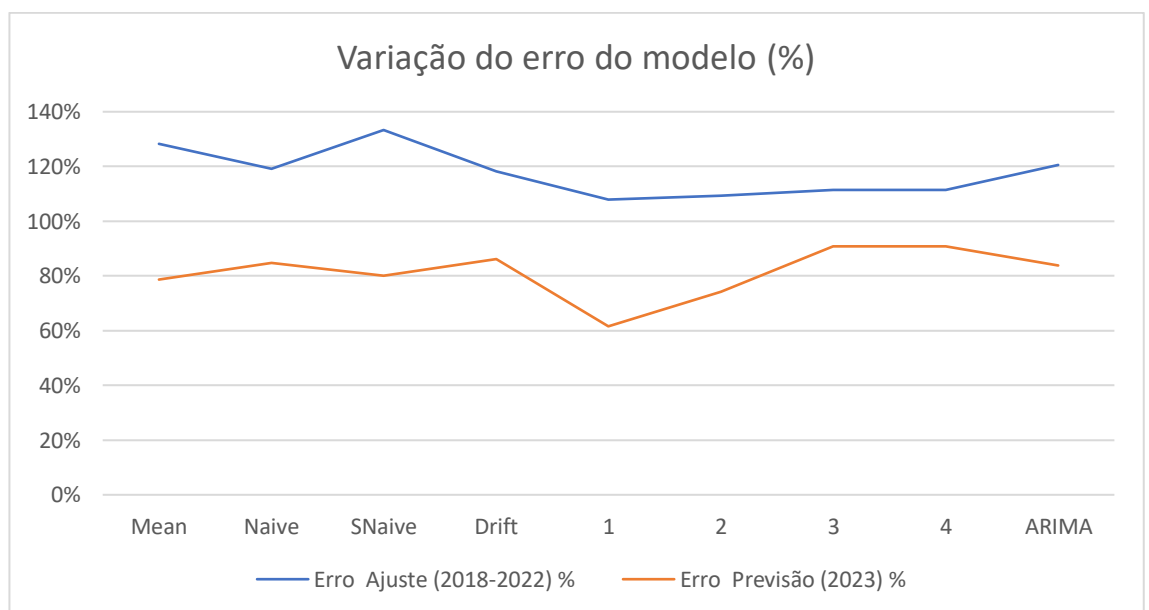
9.2.11.8. Métodos regressivos - Modelo 4



9.2.11.9. ARIMA

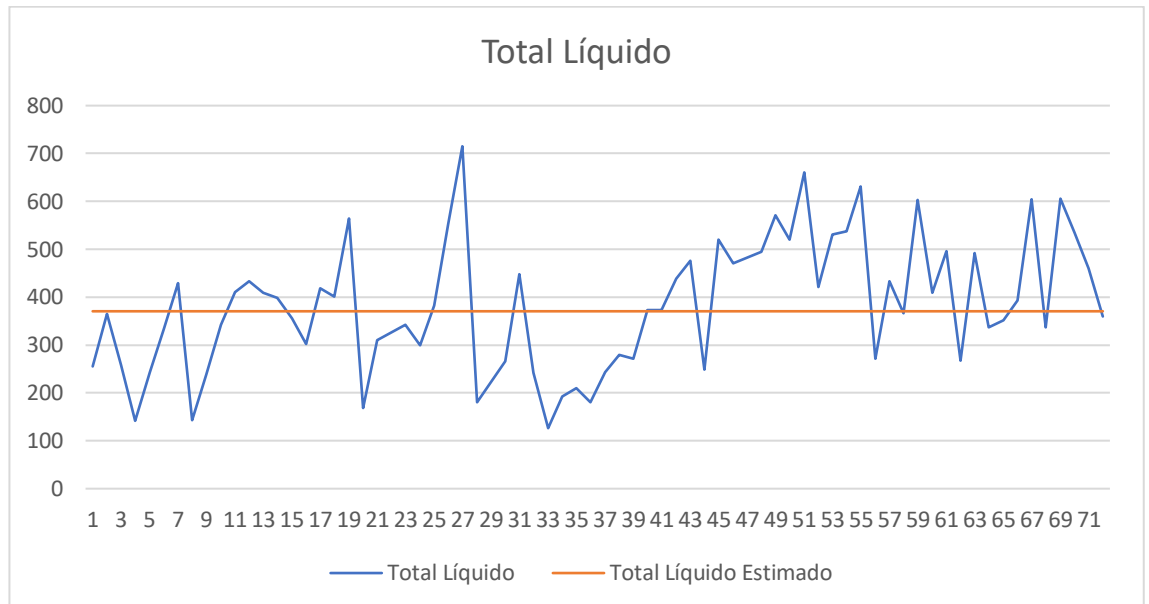


9.2.11.10. Representação gráfica dos erros dos modelos

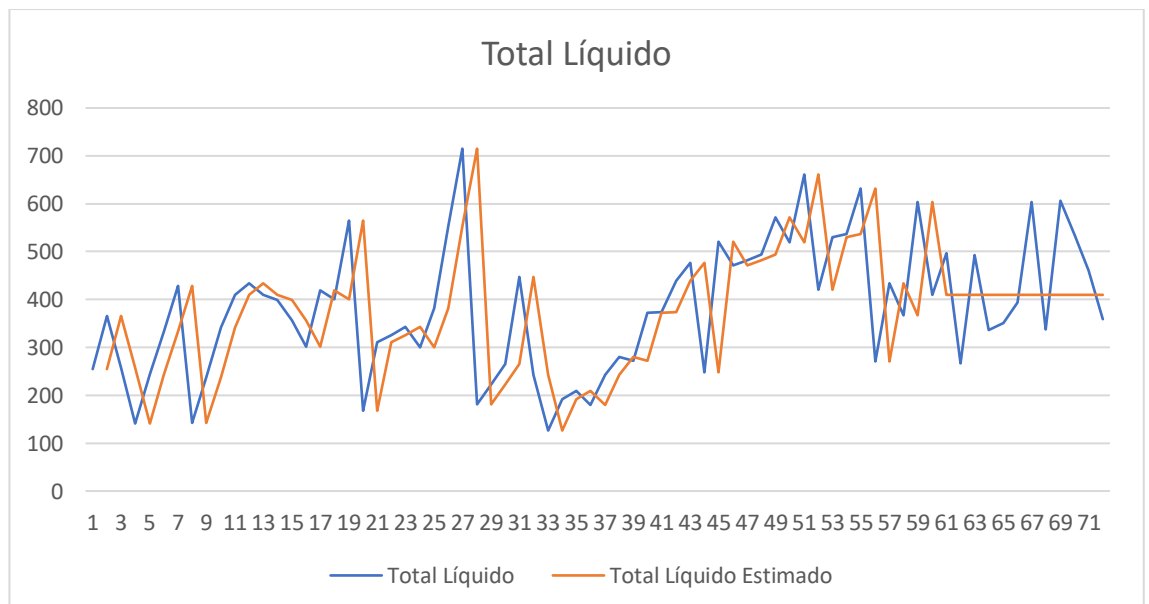


9.2.12. Total Líquido por Tipo de Peça – *Cluster 1*

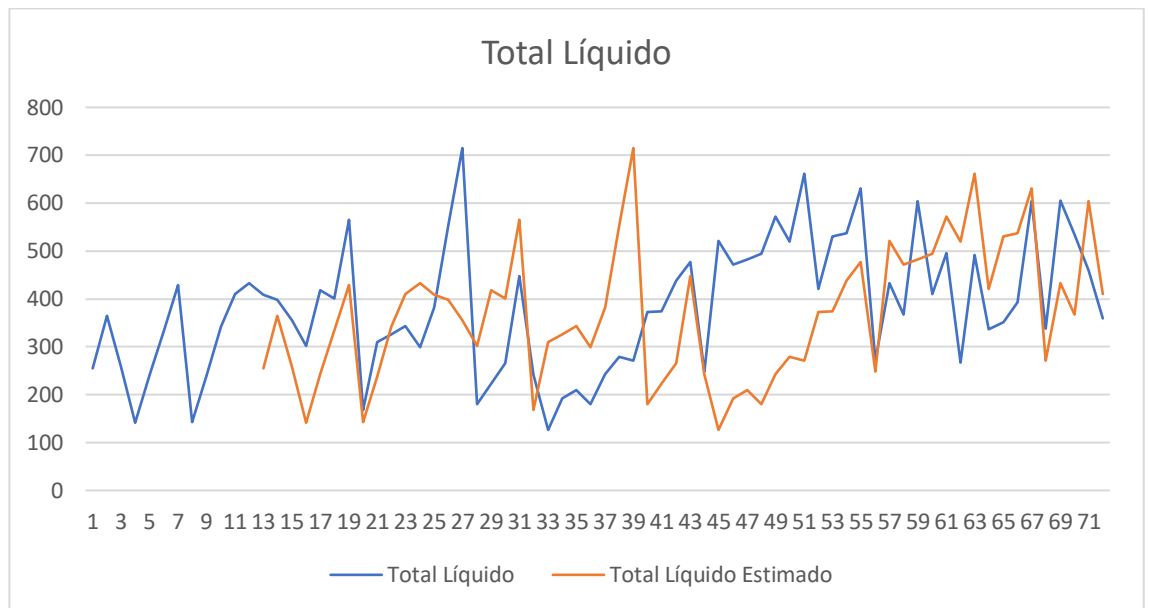
9.2.12.1. Métodos simples - Mean



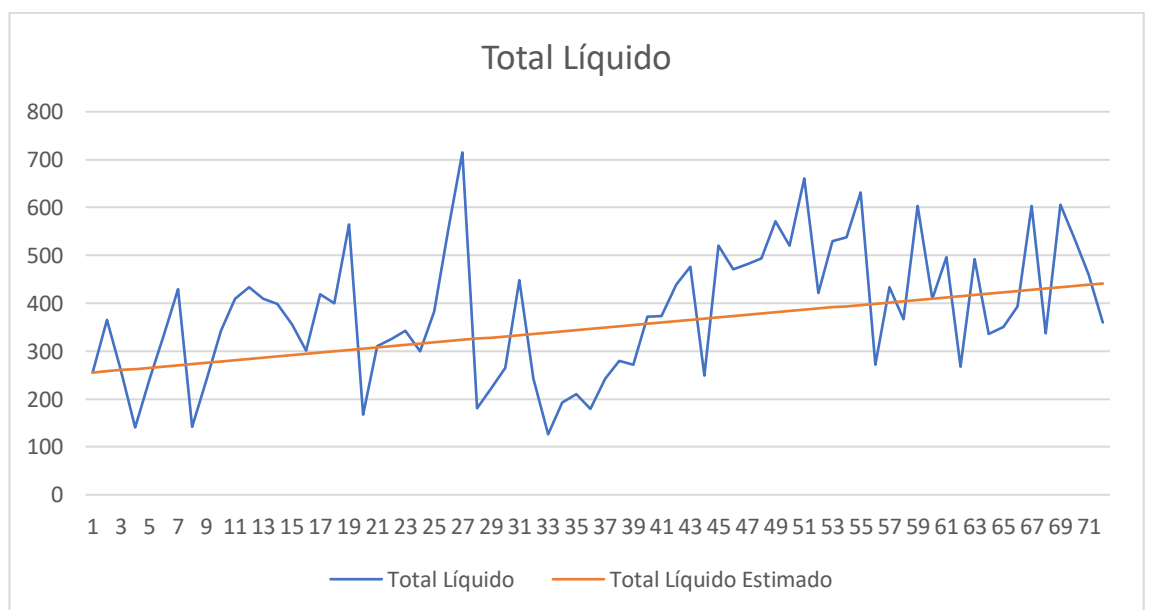
9.2.12.2. Métodos simples - Naive



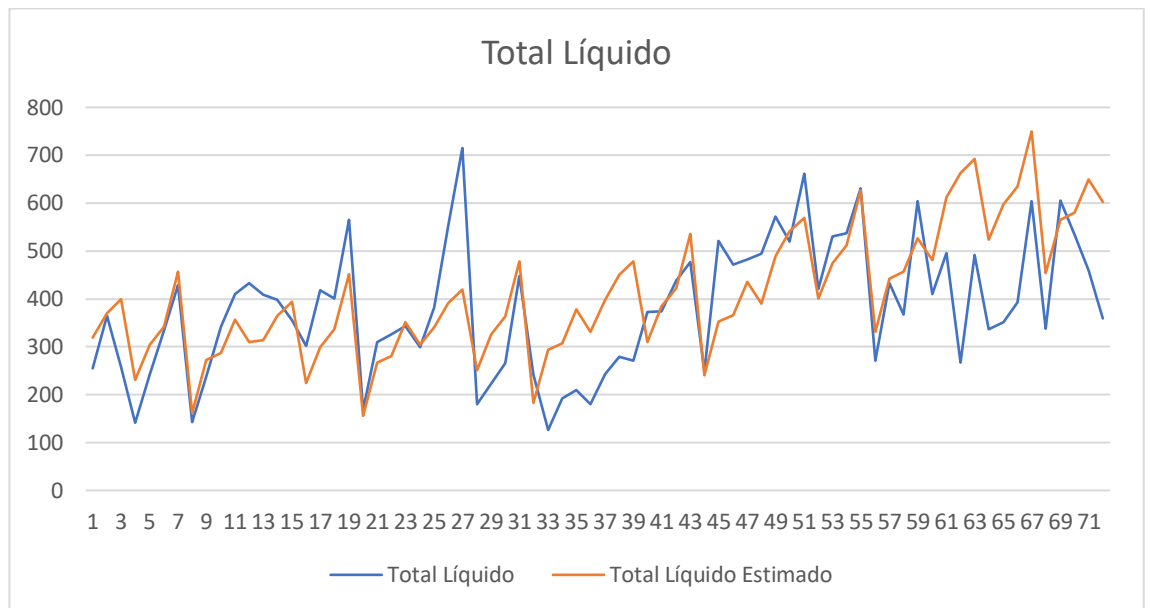
9.2.12.3. Métodos simples - SNaive



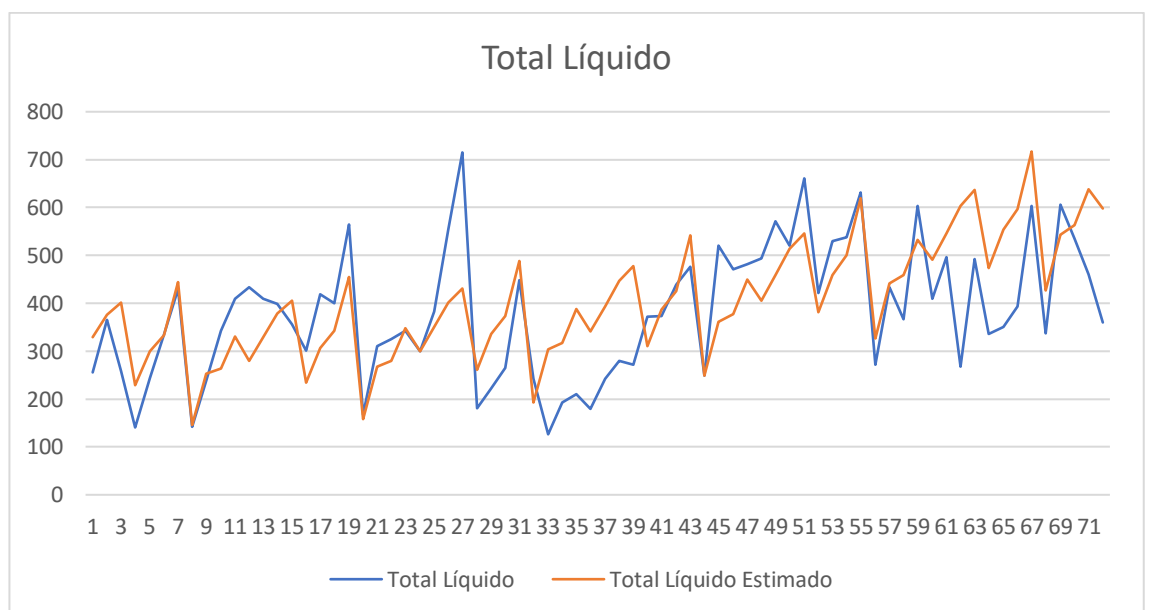
9.2.12.4. Métodos simples - Drift



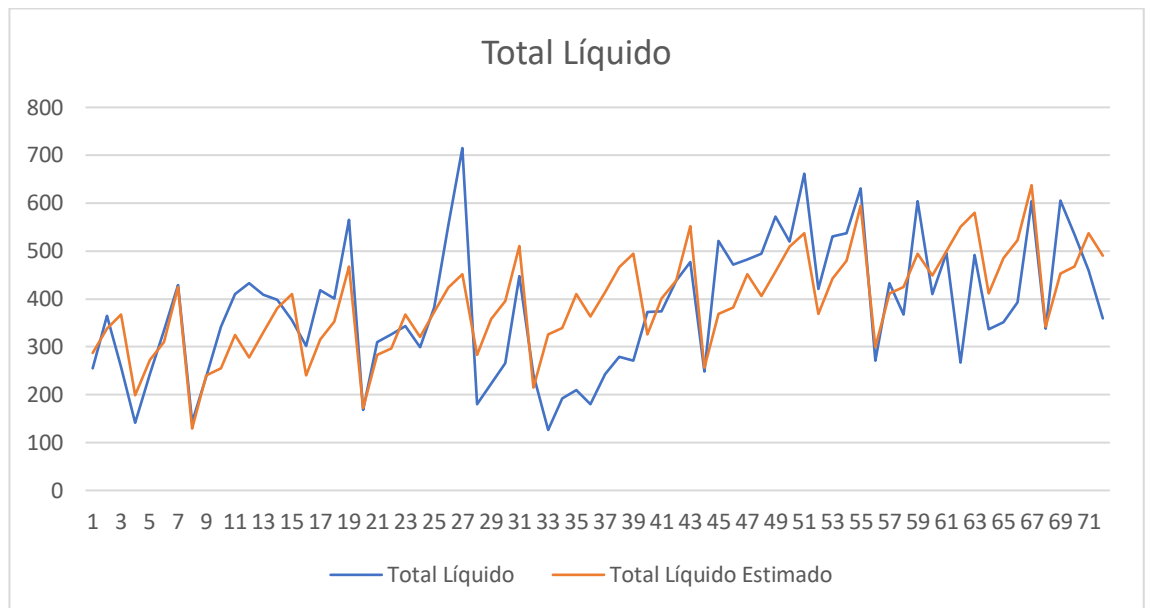
9.2.12.5. Métodos regressivos - Modelo 1



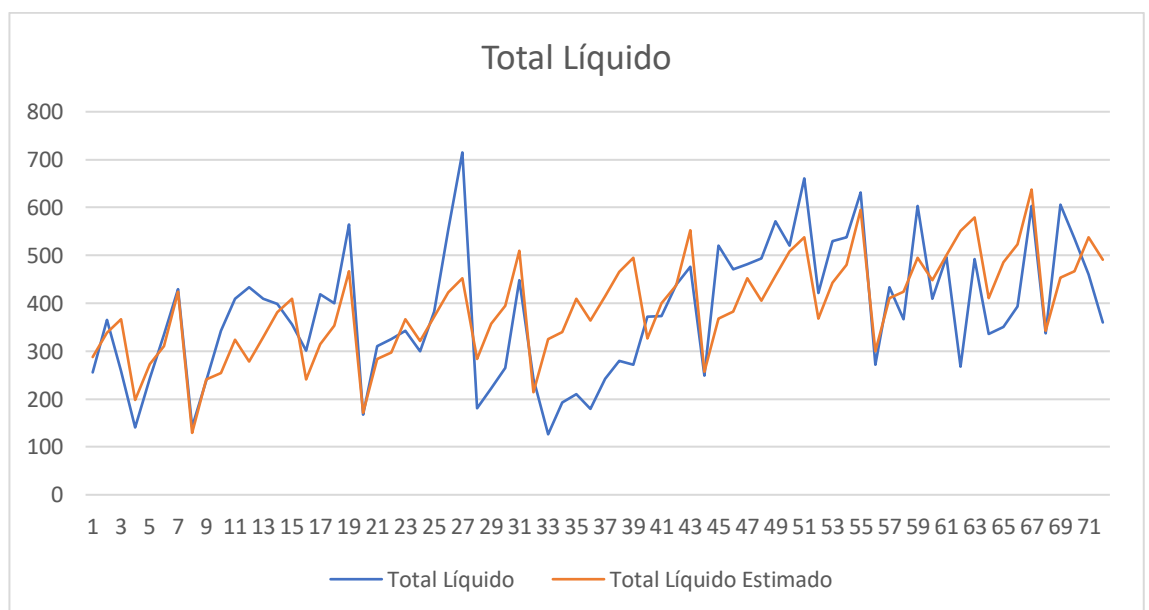
9.2.12.6. Métodos regressivos - Modelo 2



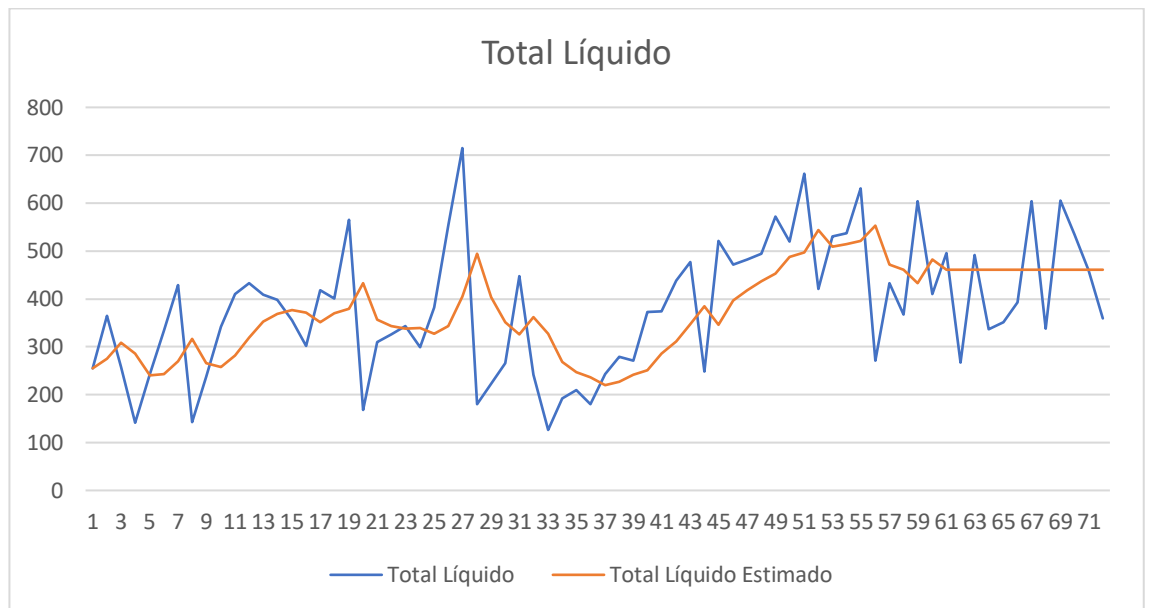
9.2.12.7. Métodos regressivos - Modelo 3



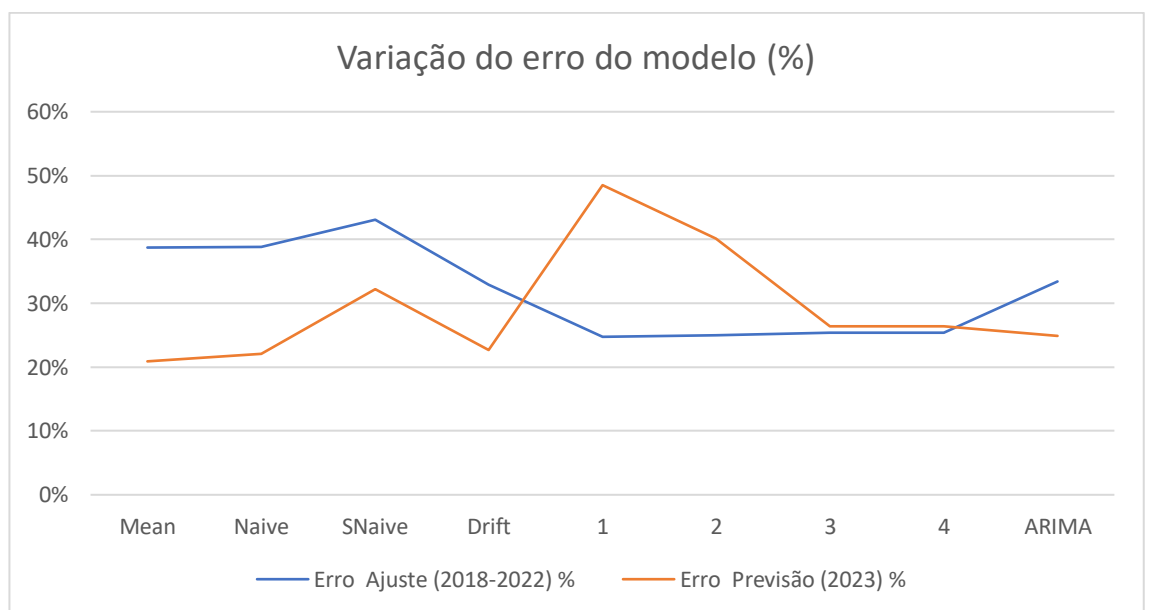
9.2.12.8. Métodos regressivos - Modelo 4



9.2.12.9. ARIMA

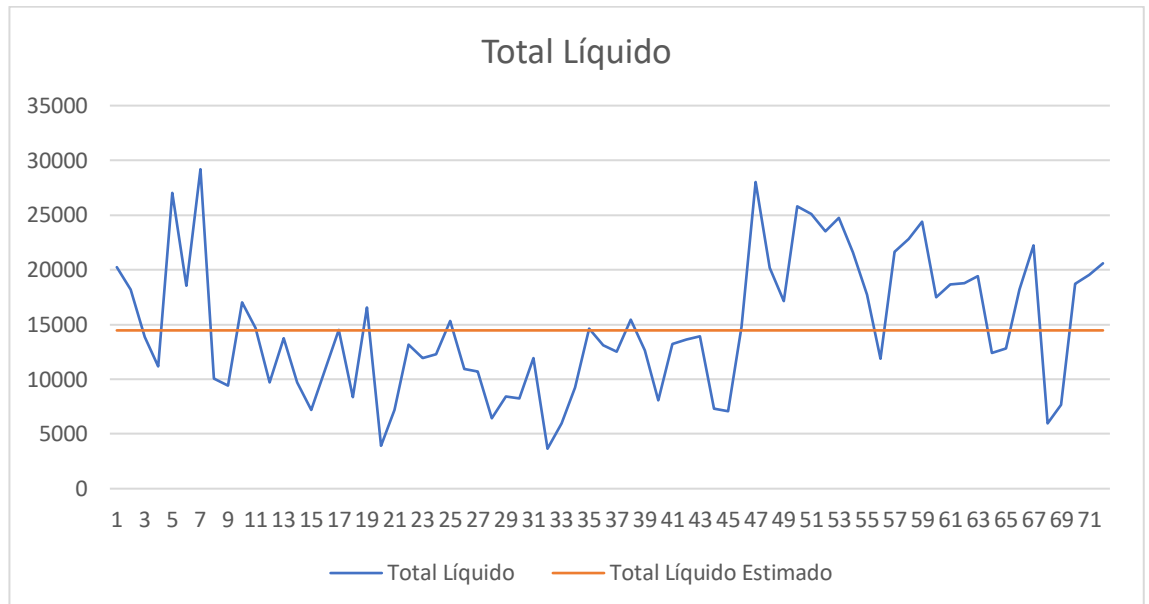


9.2.12.10. Representação gráfica dos erros dos modelos

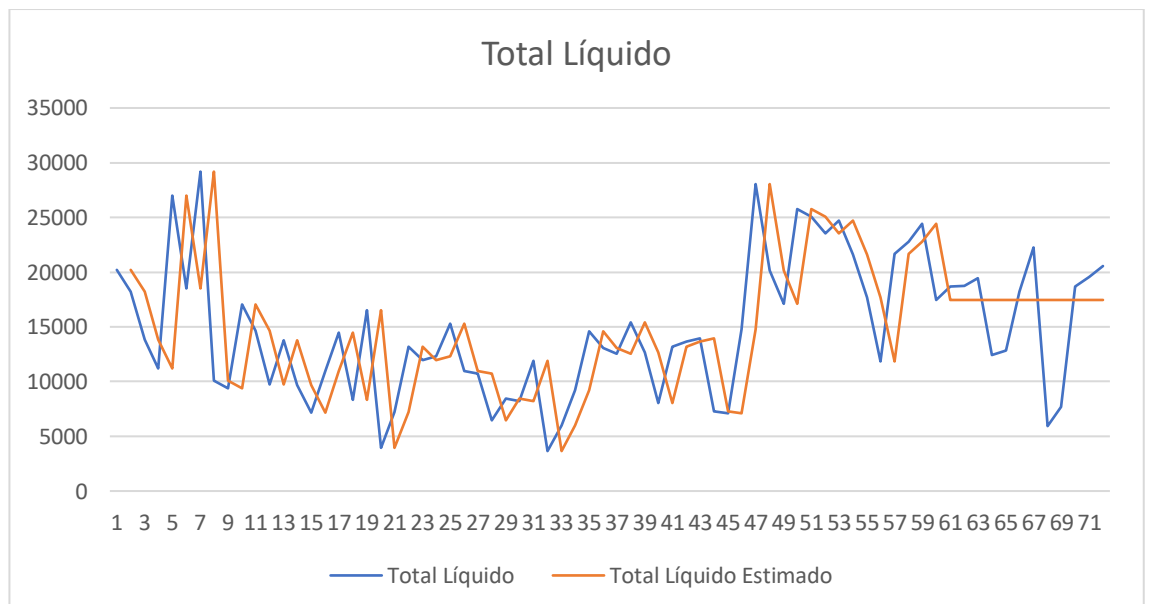


9.2.13. Total Líquido por Tipo de Peça – *Cluster 2*

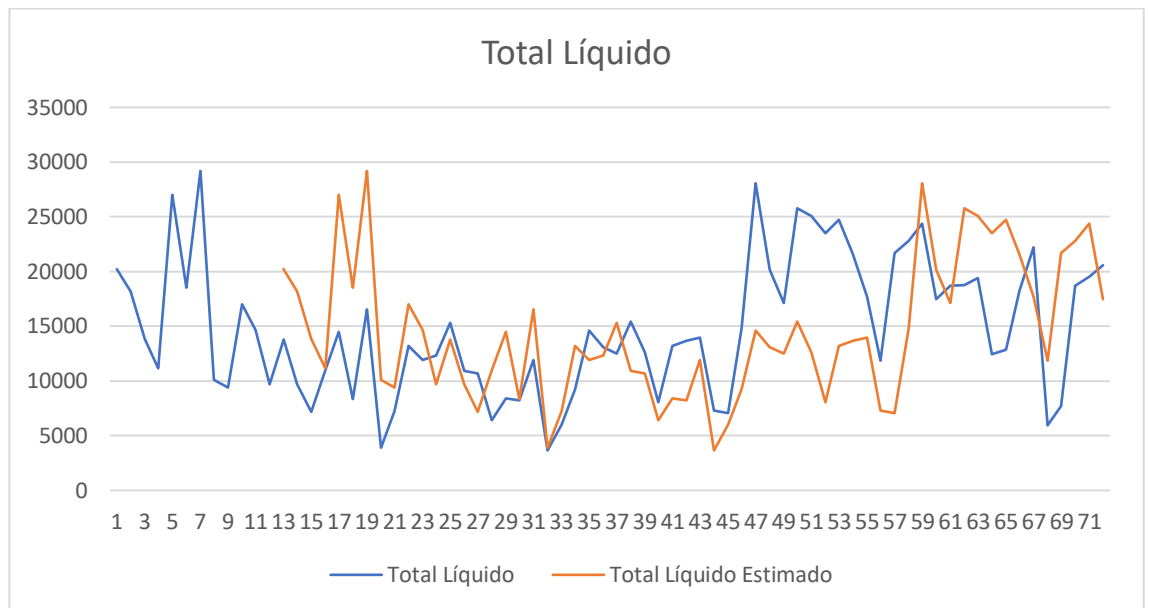
9.2.13.1. Métodos simples - Mean



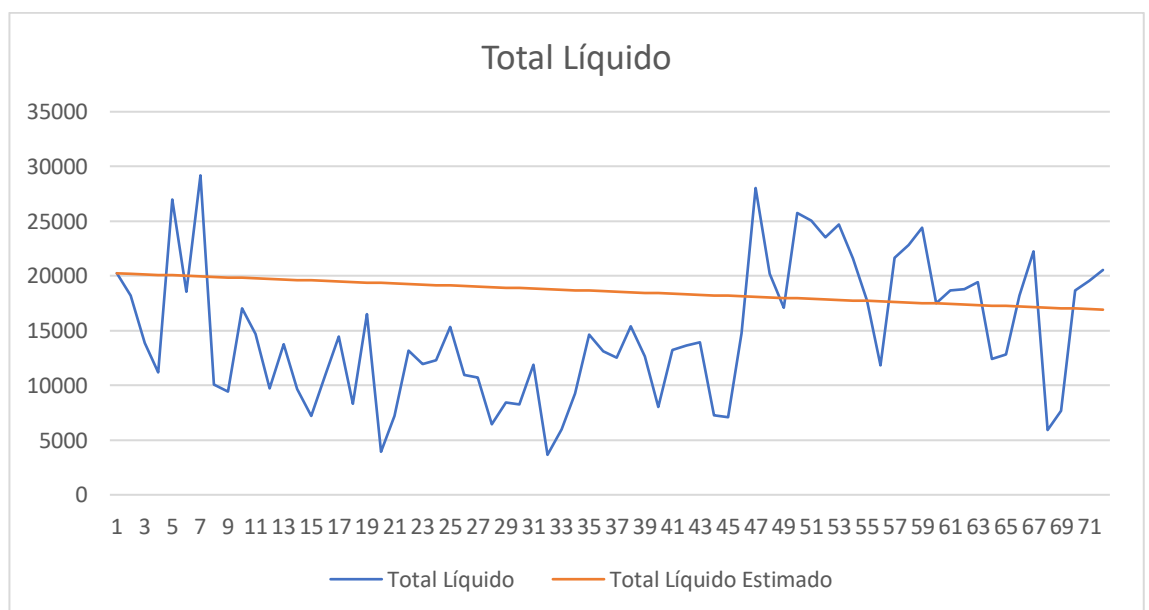
9.2.13.2. Métodos simples - Naive



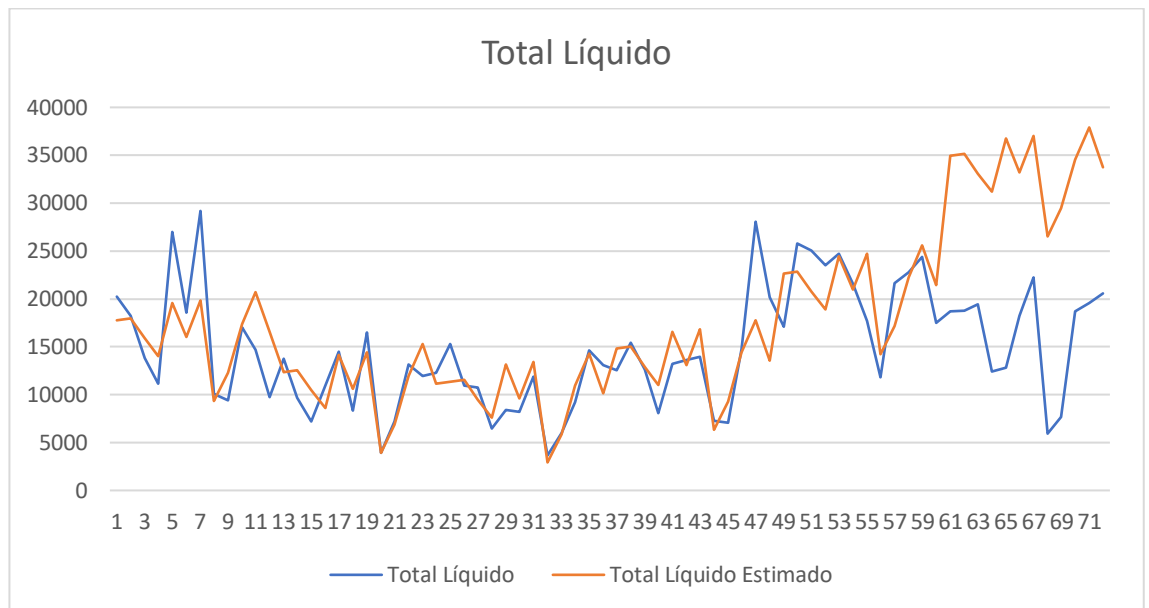
9.2.13.3. Métodos simples - SNaive



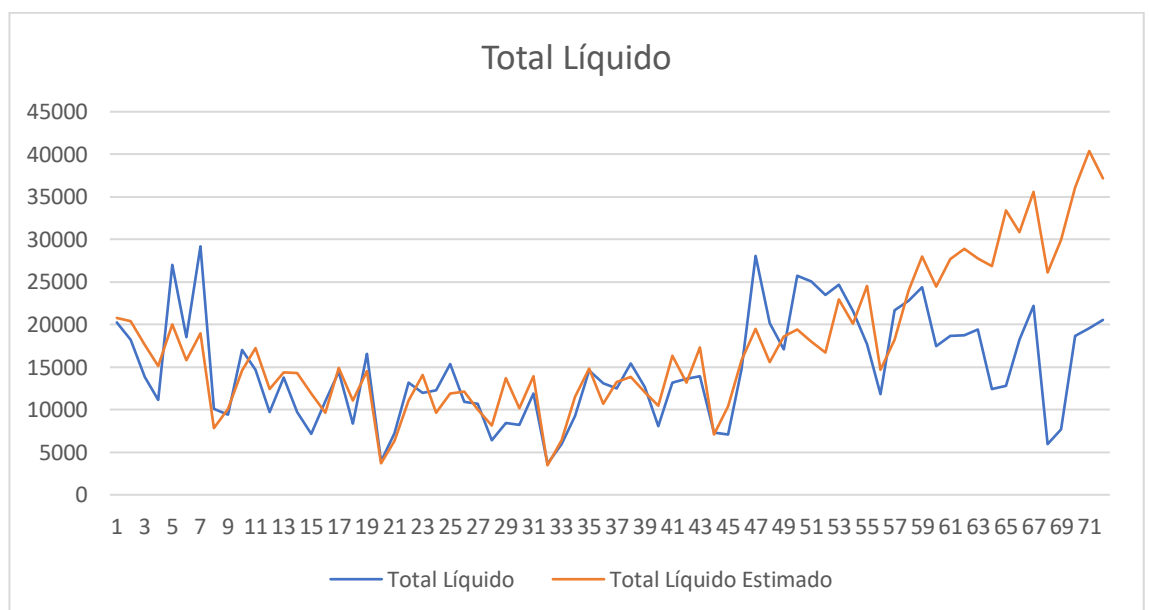
9.2.13.4. Métodos simples - Drift



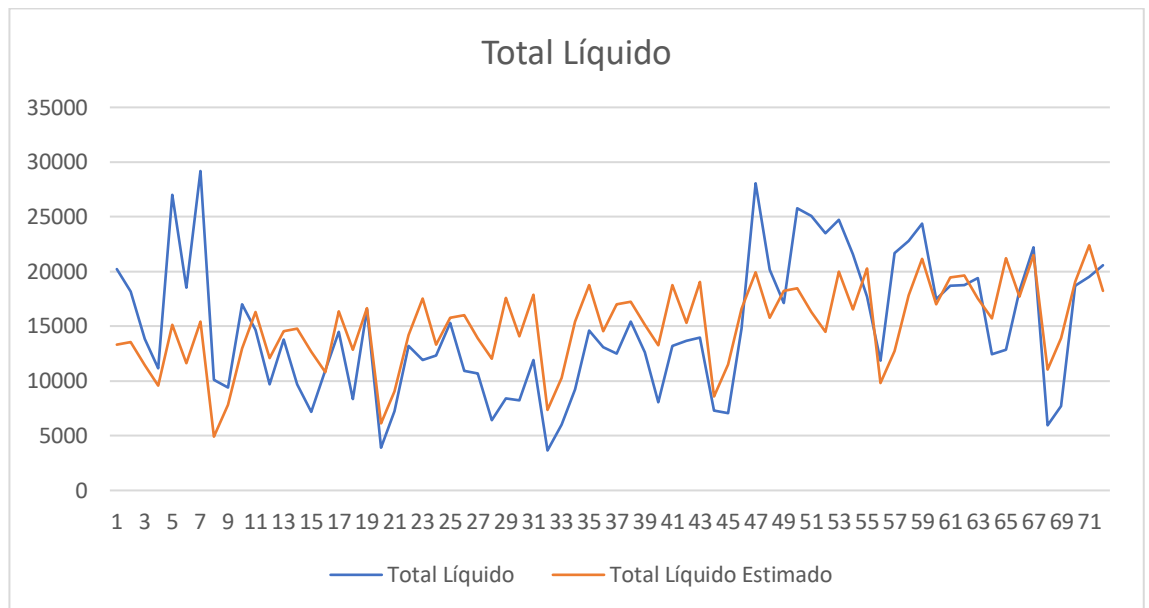
9.2.13.5. Métodos regressivos - Modelo 1



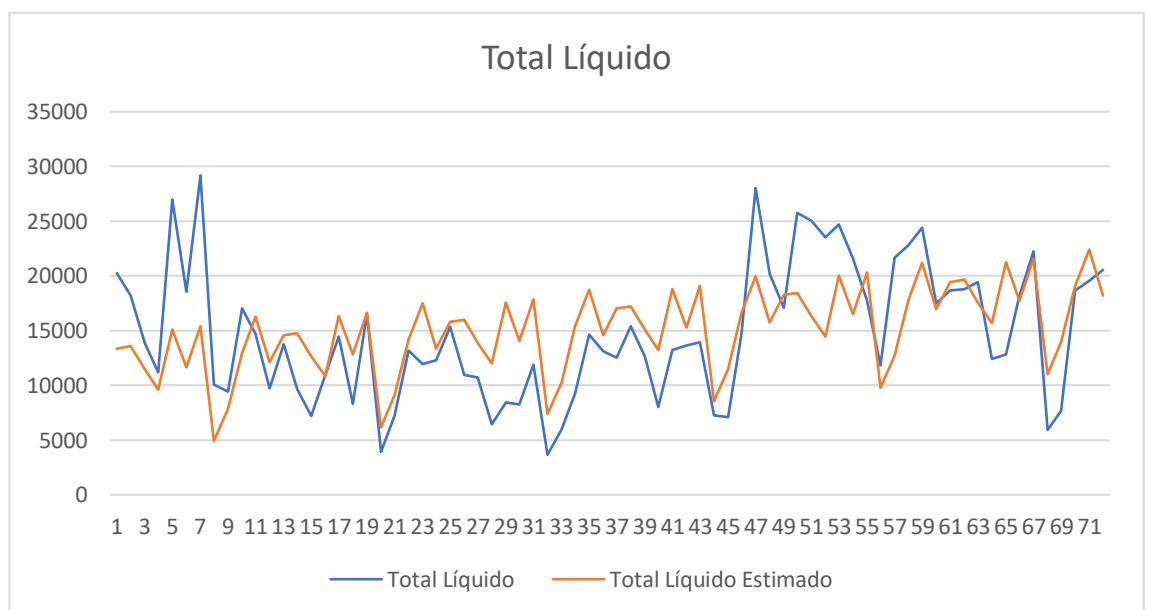
9.2.13.6. Métodos regressivos - Modelo 2



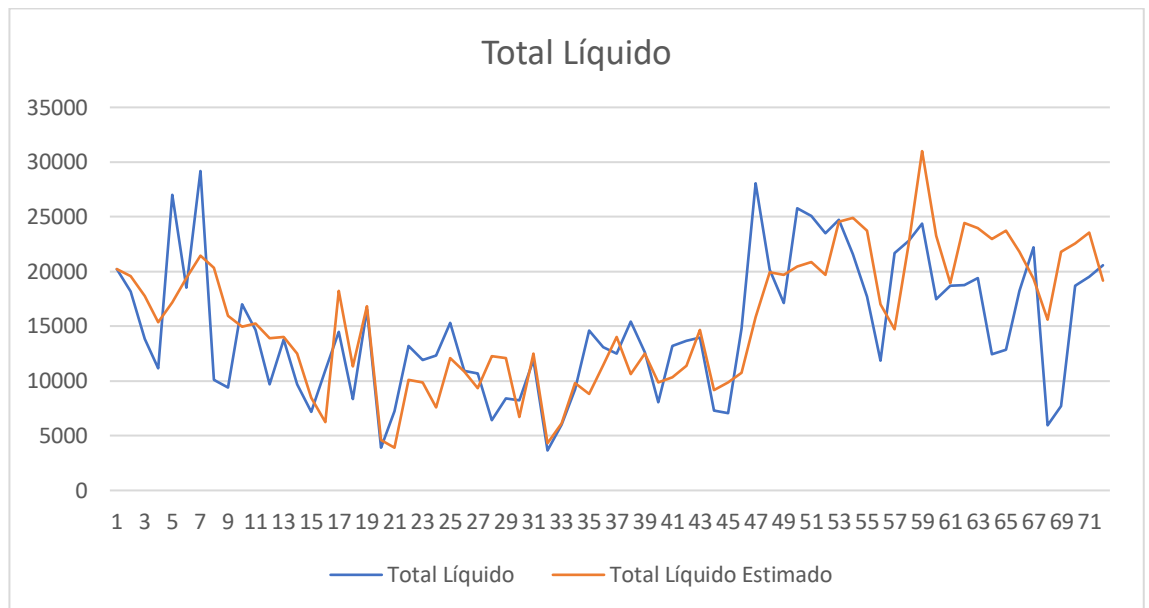
9.2.13.7. Métodos regressivos - Modelo 3



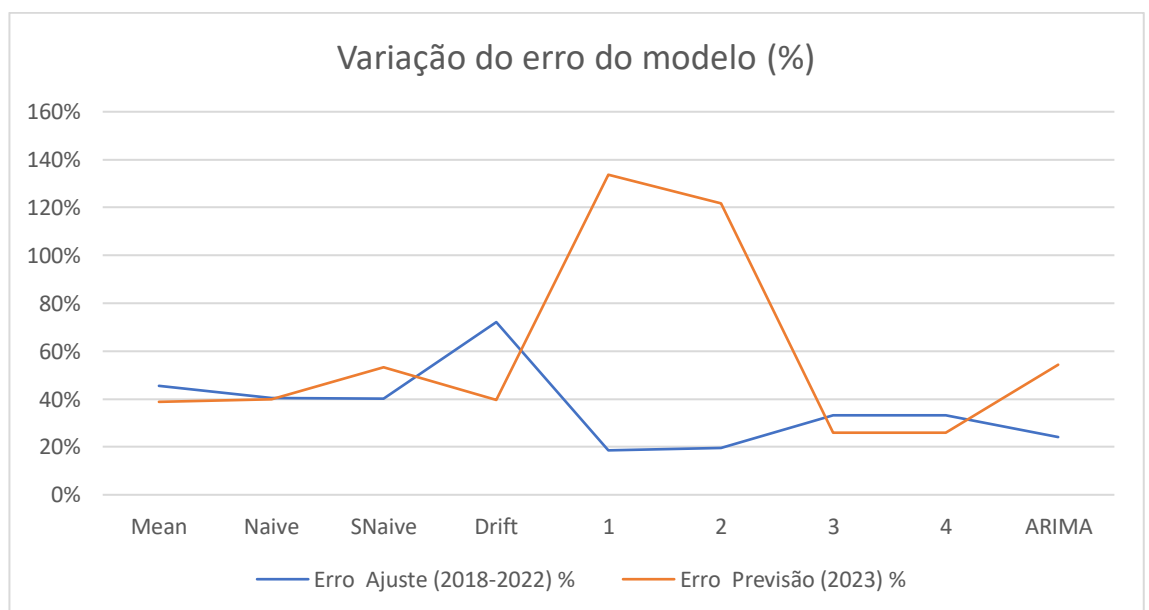
9.2.13.8. Métodos regressivos - Modelo 4



9.2.13.9. ARIMA



9.2.13.10. Representação gráfica dos erros dos modelos



9.3. Significância de coeficientes e modelos de previsão regressivos³

9.3.1. Total líquido global

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	41004,79	31009,00	20787,12	20787,12
Mês 1	1612,10	4369,79	1612,10	4369,79
Mês 2	3971,98	6646,00	3971,98	6478,97
Mês 3	3068,79	5625,73	3068,79	5325,08
Mês 4	-5883,19	-3476,74	-5883,19	-3877,59
Mês 5	2774,32	4996,88	2774,32	4529,22
Mês 6	855,24	2860,51	855,24	2359,43
Mês 7	9455,78	11210,35	9455,78	10709,27
Mês 8	-13450,93	-11980,47	-13450,93	-12448,13
Mês 9	-5861,75	-4708,79	-5861,75	-5109,65
Mês 10	-146,16	655,88	-146,16	355,24
Mês 11	6123,82	6541,55	6123,82	6374,52
Ano/Período	-14321,03	-768,15	3008,39	250,70
Ano^2/Período^2	2888,24	16,70	-	-
F	8,851	7,625	3,829	3,829

9.3.2. Total líquido por cliente – *cluster* 1

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	255,57	199,56	145,56	145,56
Mês 1	36,89	43,04	36,89	43,04
Mês 2	72,78	79,25	72,78	78,37
Mês 3	48,66	55,28	48,66	53,69
Mês 4	-40,13	-33,54	-40,13	-35,66
Mês 5	7,05	13,43	7,05	10,96
Mês 6	7,42	13,42	7,42	10,77
Mês 7	61,01	66,45	61,01	63,80
Mês 8	-71,27	-66,56	-71,27	-69,03
Mês 9	-39,02	-35,23	-39,02	-37,34
Mês 10	-26,65	-23,94	-26,65	-25,53
Mês 11	8,51	9,95	8,51	9,07
Ano/Período	-87,59	-4,82	6,71	0,56
Ano^2/Período^2	15,72	0,09	-	-
F	4,484	3,992	2,707	2,707

³ Nas tabelas os coeficientes e os F com significância encontram-se destacados a **negrito**, **itálico** e com sombreado verde.

9.3.3. Total líquido por cliente – *cluster 2*

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	5830,63	4216,52	2465,28	2465,28
Mês 1	-593,98	136,73	-593,98	136,73
Mês 2	-907,88	-214,98	-907,88	-243,60
Mês 3	-469,41	179,95	-469,41	128,45
Mês 4	-730,39	-130,29	-730,39	-198,96
Mês 5	708,70	1253,82	708,70	1173,70
Mês 6	57,64	542,06	57,64	456,22
Mês 7	1280,96	1698,95	1280,96	1613,10
Mês 8	-2298,15	-1952,32	-2298,15	-2032,44
Mês 9	-757,27	-489,31	-757,27	-557,99
Mês 10	768,46	952,83	768,46	901,32
Mês 11	1780,20	1875,24	1780,20	1846,63
Ano/Período	-2087,44	-108,12	797,14	66,43
Ano^2/Período^2	480,76	2,86	-	-
F	12,648	11,595	6,272	6,272

9.3.4. Total líquido por tipo de lavagem – *cluster 1*

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	47,73	73,17	98,12	98,12
Mês 1	14,81	9,81	14,81	9,81
Mês 2	64,18	59,22	64,18	59,63
Mês 3	37,58	32,75	37,58	33,49
Mês 4	-34,17	-38,78	-34,17	-37,80
Mês 5	10,12	5,80	10,12	6,94
Mês 6	-4,16	-8,11	-4,16	-6,89
Mês 7	36,17	32,67	36,17	33,90
Mês 8	-5,39	-8,35	-5,39	-7,21
Mês 9	1,84	-0,51	1,84	0,47
Mês 10	-25,43	-27,08	-25,43	-26,34
Mês 11	-8,16	-9,03	-8,16	-8,62
Ano/Período	37,75	2,03	-5,45	-0,45
Ano^2/Período^2	-7,20	-0,04	-	-
F	0,744	0,718	0,669	0,669

9.3.5. Total líquido por tipo de lavagem – *cluster 2*

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	10505,39	6433,01	2625,44	2625,44
Mês 1	-2287,87	841,90	-2287,87	841,90
Mês 2	-3090,01	-182,55	-3090,01	-244,77
Mês 3	-3385,28	-712,57	-3385,28	-824,56
Mês 4	-5785,92	-3360,41	-5785,92	-3509,72
Mês 5	-2383,89	-218,01	-2383,89	-392,22
Mês 6	-524,10	1369,70	-524,10	1183,05
Mês 7	-425,39	1183,88	-425,39	997,23
Mês 8	-7577,10	-6264,80	-7577,10	-6439,00
Mês 9	-5057,79	-4054,90	-5057,79	-4204,22
Mês 10	-3077,40	-2396,37	-3077,40	-2508,35
Mês 11	1844,75	2191,49	1844,75	2129,27
Ano/Período	-3339,94	-94,99	3414,30	284,52
Ano^2/Período^2	1125,71	6,22	-	-
F	26,099	21,788	14,462	14,462

9.3.6. Total líquido por tipo de lavagem – *cluster 3*

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	5987,11	3546,88	2143,00	2143,00
Mês 1	549,81	1346,36	549,81	1346,36
Mês 2	580,26	1327,33	580,26	1304,39
Mês 3	666,83	1359,84	666,83	1318,55
Mês 4	1042,45	1676,82	1042,45	1621,76
Mês 5	2136,50	2707,62	2136,50	2643,39
Mês 6	1844,17	2347,47	1844,17	2278,65
Mês 7	3782,25	4213,13	3782,25	4144,31
Mês 8	-1325,01	-971,13	-1325,01	-1035,36
Mês 9	2736,66	3008,95	2736,66	2953,90
Mês 10	2964,23	3150,34	2964,23	3109,05
Mês 11	2808,11	2903,46	2808,11	2880,52
Ano/Período	-2425,99	-67,52	868,96	72,41
Ano^2/Período^2	549,16	2,29	-	-
F	3,638	2,926	2,706	2,706

9.3.7. Total líquido por tipo de lavagem – *cluster 4*

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	8358,40	5579,39	3571,20	3571,20
Mês 1	-1847,67	-1247,10	-1847,67	-1247,10
Mês 2	-236,76	342,03	-236,76	309,22
Mês 3	795,17	1345,62	795,17	1286,55
Mês 4	-598,53	-82,99	-598,53	-161,75
Mês 5	-609,69	-135,63	-609,69	-227,50
Mês 6	-1645,77	-1219,74	-1645,77	-1318,18
Mês 7	167,16	538,59	167,16	440,15
Mês 8	-3341,79	-3031,52	-3341,79	-3123,40
Mês 9	-2269,30	-2026,75	-2269,30	-2105,51
Mês 10	152,09	320,35	152,09	261,29
Mês 11	672,40	759,81	672,40	727,00
Ano/Período	-3448,14	-145,56	655,17	54,60
Ano^2/Período^2	683,89	3,28	-	-
F	5,798	3,992	2,602	2,602

9.3.8. Total líquido por tipo de lavagem – *cluster 5*

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	2608,45	2748,32	2530,37	2530,37
Mês 1	368,68	110,17	368,68	110,17
Mês 2	1482,78	1251,34	1482,78	1247,78
Mês 3	1936,97	1731,87	1936,97	1725,46
Mês 4	-187,34	-366,80	-187,34	-375,35
Mês 5	1765,73	1611,19	1765,73	1601,22
Mês 6	474,50	344,18	474,50	333,49
Mês 7	2821,13	2714,31	2821,13	2703,63
Mês 8	433,57	349,53	433,57	339,56
Mês 9	491,87	429,91	491,87	421,36
Mês 10	596,17	555,58	596,17	549,17
Mês 11	681,62	661,68	681,62	658,12
Ano/Período	-348,94	-45,22	-282,01	-23,50
Ano^2/Período^2	11,15	0,36	-	-
F	1,184	1,199	1,310	1,310

9.3.9. Total líquido por tipo de lavagem – *cluster* 6

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	2841,29	3372,83	3407,13	3407,13
Mês 1	566,28	152,98	566,28	152,98
Mês 2	2313,51	1937,23	2313,51	1937,79
Mês 3	1061,90	722,74	1061,90	723,75
Mês 4	-136,38	-438,31	-136,38	-436,96
Mês 5	8,61	-255,97	8,61	-254,40
Mês 6	-355,66	-582,78	-355,66	-581,09
Mês 7	-208,13	-397,68	-208,13	-395,99
Mês 8	-1496,63	-1648,49	-1496,63	-1646,92
Mês 9	-1646,40	-1760,46	-1646,40	-1759,11
Mês 10	-809,32	-885,47	-809,32	-884,46
Mês 11	-250,12	-288,25	-250,12	-287,69
Ano/Período	34,14	-34,15	-450,87	-37,57
Ano^2/Período^2	-80,83	-0,06	-	-
F	4,377	4,259	4,713	4,713

9.3.10. Total líquido por tipo de lavagem – *cluster* 7

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	9334,28	7324,72	3944,30	3944,30
Mês 1	3388,73	2321,67	3388,73	2321,67
Mês 2	1969,55	1054,74	1969,55	999,50
Mês 3	1564,24	790,62	1564,24	691,20
Mês 4	808,88	165,41	808,88	32,85
Mês 5	2231,06	1706,68	2231,06	1552,02
Mês 6	1382,97	966,65	1382,97	800,95
Mês 7	2793,65	2474,33	2793,65	2308,62
Mês 8	739,64	506,28	739,64	351,62
Mês 9	426,57	268,12	426,57	135,55
Mês 10	86,65	-7,94	86,65	-107,36
Mês 11	495,48	453,71	495,48	398,48
Ano/Período	-5784,04	-433,94	-1164,06	-97,00
Ano^2/Período^2	770,00	5,52	-	-
F	5,215	6,676	2,775	2,775

9.3.11. Total líquido por tipo de lavagem – *cluster 8*

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	749,43	1052,69	1290,05	1290,05
Mês 1	681,67	716,32	681,67	716,32
Mês 2	118,37	145,99	118,37	149,87
Mês 3	-59,57	-38,20	-59,57	-31,22
Mês 4	-582,19	-566,30	-582,19	-556,99
Mês 5	-505,61	-494,42	-505,61	-483,56
Mês 6	-266,79	-259,53	-266,79	-247,89
Mês 7	54,90	59,02	54,90	70,65
Mês 8	-813,51	-811,77	-813,51	-800,91
Mês 9	-567,21	-567,07	-567,21	-557,76
Mês 10	272,06	271,38	272,06	278,36
Mês 11	-22,28	-23,01	-22,28	-19,13
Ano/Período	501,19	26,81	37,80	3,15
Ano^2/Período^2	-77,23	-0,39	-	-
F	0,748	0,719	0,738	0,738

9.3.12. Total líquido por tipo de peça – *cluster 1*

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	347,80	282,17	235,61	235,61
Mês 1	8,91	47,92	8,91	47,92
Mês 2	60,09	96,32	60,09	95,56
Mês 3	88,86	122,15	88,86	120,78
Mês 4	-79,94	-49,74	-79,94	-51,57
Mês 5	-5,67	21,28	-5,67	19,15
Mês 6	31,88	55,44	31,88	53,16
Mês 7	146,55	166,56	146,55	164,28
Mês 8	-148,87	-132,56	-148,87	-134,69
Mês 9	-37,59	-25,12	-37,59	-26,95
Mês 10	-23,71	-15,25	-23,71	-16,61
Mês 11	46,30	50,61	46,30	49,85
Ano/Período	-53,60	-1,09	42,56	3,55
Ano^2/Período^2	16,03	0,08	-	-
F	3,816	3,579	3,630	3,630

9.3.13. Total líquido por tipo de peça – *cluster 2*

Termo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Coeficientes	Coeficientes	Coeficientes	Coeficientes
Constante	26397,34	19157,96	10891,43	10891,43
Mês 1	1238,03	2357,21	1238,03	2357,21
Mês 2	1448,14	2600,65	1448,14	2465,58
Mês 3	-663,37	495,46	-663,37	252,33
Mês 4	-2525,81	-1387,68	-2525,81	-1711,86
Mês 5	3012,57	4102,99	3012,57	3724,78
Mês 6	-483,55	532,14	-483,55	126,92
Mês 7	3300,87	4214,81	3300,87	3809,59
Mês 8	-7198,25	-6413,06	-7198,25	-6791,27
Mês 9	-4283,04	-3653,63	-4283,04	-3977,81
Mês 10	849,57	1296,19	849,57	1053,05
Mês 11	4179,06	4415,88	4179,06	4280,80
Ano/Período	-12069,85	-722,21	1220,93	101,74
Ano^2/Período^2	2215,13	13,51	-	-
F	7,745	7,180	1,949	1,949

9.4. Significância de coeficientes e do modelo ARIMA⁴

9.4.1. Total líquido global

Modelo ARIMA (0,1,1)(0,1,1)[12]		
Componente	Termo	Coefficiente
Não sazonal	Média móvel 1 (MA1)	-0,47348
Sazonal	Média móvel 1 (SMA1)	-0,52549

Teste Box-Ljung do Modelo ARIMA - Qui-quadrado	
Valor	p
0,25437	0,614

9.4.2. Total líquido por cliente – *cluster* 1

Modelo ARIMA (1,0,0)(1,0,0)[12]		
Componente	Termo	Coefficiente
Não sazonal	Autoregressivo 1 (AR1)	0,48184
Sazonal	Autoregressivo 1 (SAR1)	0,29991

Teste Box-Ljung do Modelo ARIMA - Qui-quadrado	
Valor	p
0,15350	0,6952

9.4.3. Total líquido por cliente – *cluster* 2

Modelo ARIMA (4,1,0)(1,0,0)[12]		
Componente	Termo	Coefficiente
Não sazonal	Autoregressivo 1 (AR1)	-0,58811
	Autoregressivo 2 (AR2)	-0,49539
	Autoregressivo 3 (AR3)	-0,51221
	Autoregressivo 4 (AR4)	-0,26417
Sazonal	Autoregressivo 1 (SAR1)	0,64801

Teste Box-Ljung do Modelo ARIMA - Qui-quadrado	
Valor	p
0,12813	0,7204

⁴ Nas tabelas os coeficientes com significância encontram-se destacados a **negrito, *itálico*** e com sombreado verde. Os valores do teste de Box-Ljung destacados a **negrito, *itálico*** e com sombreado verde, significam que o modelo é adequado, pois o valor de p é superior a 0,05.

9.4.4. Total líquido por tipo de lavagem – *Cluster 1*

Modelo ARIMA (0,0,0)		
Componente	Termo	Coeficiente
Teste Box-Ljung do Modelo ARIMA - Qui-quadrado		
Valor	p	
1,23760	0,2659	

9.4.5. Total líquido por tipo de lavagem – *Cluster 2*

Modelo ARIMA (0,1,1)(1,0,0)[12]		
Componente	Termo	Coeficiente
Não sazonal	Média móvel 1 (MA1)	-0,64702
Sazonal	Autoregressivo 1 (SAR1)	0,49585
Teste Box-Ljung do Modelo ARIMA - Qui-quadrado		
Valor	p	
0,67326	0,4119	

9.4.6. Total líquido por tipo de lavagem – *Cluster 3*

Modelo ARIMA (0,1,1)(1,0,0)[12]		
Componente	Termo	Coeficiente
Não sazonal	Média móvel 1 (MA1)	-0,59617
Sazonal	Autoregressivo 1 (SAR1)	0,42559
Teste Box-Ljung do Modelo ARIMA - Qui-quadrado		
Valor	p	
0,07467	0,7847	

9.4.7. Total líquido por tipo de lavagem – *Cluster 4*

Modelo ARIMA (0,1,1)(1,0,0)[12]		
Componente	Termo	Coeficiente
Não sazonal	Média móvel 1 (MA1)	-0,50814
Sazonal	Autoregressivo 1 (SAR1)	0,52192

Teste Box-Ljung do Modelo ARIMA - Qui-quadrado	
Valor	p
0,16649	0,6847

9.4.8. Total líquido por tipo de lavagem – *Cluster 5*

Modelo ARIMA (1,0,0)		
Componente	Termo	Coefficiente
Não sazonal	Autoregressivo 1 (AR1)	0,23809

Teste Box-Ljung do Modelo ARIMA - Qui-quadrado	
Valor	p
0,07419	0,7853

9.4.9. Total líquido por tipo de lavagem – *Cluster 6*

Modelo ARIMA (0,1,0)		
Componente	Termo	Coefficiente

9.4.10. Total líquido por tipo de lavagem – *Cluster 7*

Modelo ARIMA (0,1,1)		
Componente	Termo	Coefficiente
Não sazonal	Média móvel 1 (MA1)	-0,64715

Teste Box-Ljung do Modelo ARIMA - Qui-quadrado	
Valor	p
0,04882	0,8251

9.4.11. Total líquido por tipo de lavagem – *Cluster 8*

Modelo ARIMA (2,0,0)		
Componente	Termo	Coefficiente
Não sazonal	Autoregressivo 1 (AR1)	0,20537
	Autoregressivo 2 (AR2)	0,23146

Teste Box-Ljung do Modelo ARIMA - Qui-quadrado	
Valor	p
0,06920	0,7925

9.4.12. Total líquido por tipo de peça – *Cluster 1*

Modelo ARIMA (0,1,1)		
Componente	Termo	Coefficiente
Não sazonal	Média móvel 1 (MA1)	-0,71106

Teste Box-Ljung do Modelo ARIMA - Qui-quadrado	
Valor	p
0,06007	0,8064

9.4.13. Total líquido por tipo de peça – *Cluster 2*

Modelo ARIMA (0,1,1)(1,0,0)[12]		
Componente	Termo	Coefficiente
Não sazonal	Média móvel 1 (MA1)	-0,58534
Sazonal	Autoregressivo 1 (SAR1)	0,63228

Teste Box-Ljung do Modelo ARIMA - Qui-quadrado	
Valor	p
0,59787	0,4394

9.5. Resíduos dos modelos de previsão

9.5.1. Total líquido global

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	30950,52	928,22	-	-	-	-233,57	-3676,82	5542,91	5542,91	17,87
2	33541,77	3519,47	2591,25	-	2527,96	-2,20	-2643,74	5774,28	5774,28	10,00
3	24675,33	-5346,97	-8866,44	-	-6401,76	-7965,45	-9805,27	-2188,97	-2188,97	-0,95
4	17107,59	-12914,71	-7567,74	-	-14032,79	-6581,21	-7619,32	-804,73	-804,73	-7,38
5	37213,86	7191,56	20106,27	-	6010,20	4867,55	4631,16	10644,03	10644,03	12,30
6	32562,91	2540,61	-4650,95	-	1295,96	2135,68	2701,01	7912,16	7912,16	5,96
7	47202,52	17180,22	14639,61	-	15872,29	8174,75	9541,79	13951,23	13951,23	18,73
8	16061,4	-13960,90	-31141,12	-	-15332,12	-59,66	2109,10	5716,82	5716,82	-12,87
9	19386,85	-10635,45	3325,45	-	-12069,95	-4323,39	-1352,92	1453,09	1453,09	-8,29
10	31390,66	1368,36	12003,81	-	-129,43	1964,83	5737,02	7741,31	7741,31	3,96
11	31880,71	1858,41	490,05	-	297,34	-3815,10	758,80	1961,37	1961,37	4,06
12	27916,21	-2106,09	-3964,50	-	-3730,45	-1655,78	3719,84	4120,70	4120,70	-16,88
13	30959,03	936,73	3042,82	8,51	-750,91	5431,26	2743,45	2543,02	2543,02	-101,39
14	26425,59	-3596,71	-4533,44	-7116,18	-5347,64	-1462,06	-3749,01	-4350,30	-4350,30	-5700,45
15	22119,81	-7902,49	-4305,78	-2555,52	-9716,70	-4864,65	-6750,74	-7752,89	-7752,89	1565,91
16	23627,86	-6394,44	1508,05	6520,27	-8271,94	5595,38	4110,14	2707,14	2707,14	8722,23
17	32056,43	2034,13	8428,57	-5157,43	93,35	5366,44	4282,06	2478,20	2478,20	-6219,55
18	25168,8	-4853,50	-6887,63	-7394,11	-6857,57	397,89	-285,63	-2490,35	-2490,35	-4920,84
19	40246,68	10224,38	15077,88	-6955,84	8157,03	6875,23	6592,57	3986,99	3986,99	-1941,37
20	10971,2	-19051,10	-29275,48	-5090,20	-21181,74	506,46	624,66	-2381,78	-2381,78	732,29
21	20249,63	-9772,67	9278,43	862,78	-11966,59	2195,71	2714,76	-692,53	-692,53	5616,35
22	26855,55	-3166,75	6605,92	-4535,11	-5423,96	3086,04	4005,95	197,80	197,80	-2119,12
23	26343,92	-3678,38	-511,63	-5536,79	-5998,87	-3695,58	-2374,81	-6583,81	-6583,81	-1890,11
24	24887,07	-5135,23	-1456,85	-3029,14	-7519,01	971,40	2693,03	-1916,84	-1916,84	1324,84
25	31374,88	1352,58	6487,81	415,85	-1094,48	5726,96	4760,77	-49,52	-49,52	4493,37
26	34188,45	4166,15	2813,57	7762,86	1655,80	6180,65	5214,46	404,17	404,17	6169,68
27	40744,6	10722,30	6556,15	18624,79	8148,67	13639,98	12673,80	7863,51	7863,51	15202,34
28	14023,33	-15998,97	-26721,27	-9604,53	-18635,89	-4129,31	-5095,49	-9905,78	-9905,78	-16605,11
29	17775,97	-12246,33	3752,64	-14280,46	-14946,53	-9034,18	-10000,36	-14810,65	-14810,65	-17066,86
30	19381,84	-10640,46	1605,87	-5786,96	-13403,95	-5509,22	-6475,41	-11285,70	-11285,70	-723,57
31	30718,79	696,49	11336,95	-9527,89	-2130,28	-2772,81	-3739,00	-8549,29	-8549,29	-3801,19
32	13823,3	-16199,00	-16895,49	2852,10	-19089,06	3238,40	2272,22	-2538,07	-2538,07	10972,36
33	11292,94	-18729,36	-2530,36	-8956,69	-21682,70	-6881,14	-7847,32	-12657,61	-12657,61	-3894,64
34	17309,53	-12712,77	6016,59	-9546,02	-15729,40	-6580,14	-7546,32	-12356,61	-12356,61	-4575,71
35	23416,82	-6605,48	6107,29	-2927,10	-9685,39	-6742,83	-7709,01	-12519,31	-12519,31	3862,40
36	20650,1	-9372,20	-2766,72	-4236,97	-12515,40	-3385,73	-4351,91	-9162,20	-9162,20	1559,42
37	22732,8	-7289,50	2082,70	-8642,08	-10495,98	-8811,76	-7090,13	-11699,99	-11699,99	-2010,12
38	27156,85	-2865,45	4424,05	-7031,60	-6135,22	-6747,59	-5426,82	-9635,82	-9635,82	2823,19
39	24034,81	-5987,49	-3122,04	-16709,79	-9320,54	-8966,44	-8046,53	-11854,68	-11854,68	-2001,57
40	23692,41	-6329,89	-342,40	9669,08	-9726,23	-356,86	162,19	-3245,10	-3245,10	13175,26
41	28896,32	-1125,98	5203,91	11120,35	-4585,60	-3810,46	-3692,26	-6698,70	-6698,70	3010,18
42	32068,75	2046,45	3172,43	12686,91	-1476,46	1281,05	998,39	-1607,18	-1607,18	6705,66
43	33946,06	3923,76	1877,31	3227,27	337,57	-5442,18	-6125,70	-8330,41	-8330,41	-7963,78
44	17715,12	-12307,18	-16230,94	3891,82	-15956,66	1233,59	149,21	-1654,65	-1654,65	3358,30
45	28949,58	-1072,72	11234,46	17656,64	-4785,48	4878,87	3393,63	1990,63	1990,63	10644,65
46	34562,05	4539,75	5612,47	17252,52	763,70	4775,75	2889,66	1887,51	1887,51	3255,33
47	48293,25	18270,95	13731,20	24876,43	14431,62	12236,97	9950,01	9348,73	9348,73	12160,44
48	40923,8	10901,50	-7369,45	20273,70	6998,88	10991,34	8303,53	8103,10	8103,10	1053,98
49	41104,77	11082,47	180,97	18371,97	7116,57	-2112,90	3262,72	3663,58	3663,58	-2831,63
50	47608,74	17586,44	6503,97	20451,89	13557,25	2031,19	6605,10	7807,67	7807,67	2699,52
51	52830,91	22808,61	5222,17	28796,10	18716,14	8156,55	11928,74	13933,03	13933,03	7866,42
52	41194,38	11172,08	-11636,53	17501,97	7016,32	5472,00	8442,47	11248,48	11248,48	-196,77
53	46990,54	16968,24	5796,16	18094,22	12749,20	2610,65	4779,41	8387,13	8387,13	-1181,21
54	44155,4	14133,10	-2835,14	12086,65	9850,77	1694,59	3061,63	7471,07	7471,07	-3806,44
55	44226,35	14204,05	70,95	10280,29	9858,44	-6835,00	-6269,67	-1058,52	-1058,52	-9400,87
56	23235,84	-6786,46	-20990,51	5520,72	-11195,36	-4918,80	-5155,19	857,68	857,68	-5491,08
57	39873,77	9851,47	16637,93	10924,19	5379,29	4129,95	3091,84	9906,43	9906,43	7503,69
58	38212,92	8190,62	-1660,85	3650,87	3655,15	-3246,49	-5086,31	2529,99	2529,99	-4633,79
59	49745,94	19723,64	11533,02	1452,69	15124,89	2016,55	-624,99	7793,02	7793,02	1138,10
60	34684,34	4662,04	-15061,60	-6239,46	0,00	-6921,23	-10364,48	-1144,75	-1144,75	-9584,64
xbarra		0,00	63,29	3207,78	-2842,50	0,00	0,00	0,00	0,00	283,01
61	39528,21	9505,91	4843,87	-1576,56	4780,58	-21139,04	-11143,25	-921,37	-921,37	-1635,96
62	29986,81	-35,49	-4697,53	-17621,93	-4824,10	-33040,32	-24247,10	-12822,65	-12822,65	-15564,36
63	40084,20	10061,90	5399,86	-12746,71	5210,00	-22039,74	-14449,10	-1822,08	-1822,08	-7232,81
64	26555,30	-3467,00	-8129,04	-14639,08	-8382,18	-26616,66	-20228,59	-6399,00	-6399,00	-11181,07
65	27563,96	-2458,34	-7120,38	-19426,58	-7436,81	-34265,51	-29080,02	-14047,85	-14047,85	-16539,13
66	34737,75	4715,45	53,41	-9417,65	-326,30	-25172,64	-21189,72	-4954,97	-4954,97	-8232,97
67	47594,44	17572,14	12910,10	3368,09	12467,10	-20916,49	-18136,14	-698,82	-698,82	568,17
68	20108,39	-9913,91	-14575,95	-3127,45	-15082,23	-25495,83	-23918,06	-5278,17	-5278,17	-6472,71
69	33026,53	3004,23	-1657,81	-6847,24	-2227,38	-20166,87	-19791,67	50,79	50,79	-4894,31
70	41174,18	11151,88	6489,84	2961,26	5856,99	-17734,81	-18562,18	2482,85	2482,85	621,23
71	38876,34	8854,04	4192,00	-10869,60	3495,86	-26302,64	-28332,58	-6084,97	-6084,97	-11462,32
72	35658,52	5636,22	974,18	974,18	214,76	-23396,63	-26629,15	-3178,97	-3178,97	-4927,16

9.5.2. Total líquido por cliente – *cluster* 1

Período	Observado	Resíduos								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	234,6102174	63,50	-	-	-	14,02	-3,25	45,46	45,46	49,70
2	254,2421739	83,13	19,63	-	21,25	-2,23	-15,27	29,20	29,20	48,12
3	187,6105435	16,50	-66,63	-	-43,76	-44,75	-53,55	-13,32	-13,32	-24,48
4	86,51271739	-84,60	-101,10	-	-143,24	-57,05	-61,62	-25,62	-25,62	-90,30
5	257,461087	86,35	170,95	-	29,33	66,72	66,39	98,15	98,15	119,23
6	205,7943478	34,68	-51,67	-	-20,72	14,68	18,58	46,11	46,11	-8,68
7	308,7680435	137,65	102,97	-	83,88	64,06	72,20	95,49	95,49	113,22
8	96,75108696	-74,36	-212,02	-	-126,52	-15,68	-3,31	15,75	15,75	-136,55
9	103,7777174	-67,34	7,03	-	-117,88	-40,90	-24,29	-9,46	-9,46	-32,76
10	158,3627174	-12,75	54,59	-	-61,67	1,31	22,16	32,75	32,75	15,31
11	176,2344565	5,12	17,87	-	-42,18	-15,98	9,10	15,46	15,46	5,70
12	191,6166304	20,50	15,38	-	-25,18	7,92	37,24	39,35	39,35	8,97
13	227,0884783	55,98	35,47	-7,52	11,91	46,94	32,29	31,23	31,23	30,26
14	174,6236957	3,51	-52,46	-79,62	-38,93	-41,41	-53,95	-57,12	-57,12	-40,68
15	139,6804348	-31,43	-34,94	-47,93	-72,26	-52,24	-62,66	-67,95	-67,95	-27,52
16	151,878913	-19,23	12,20	65,37	-58,44	48,76	40,45	33,04	33,04	22,20
17	178,1330435	7,02	26,25	-79,33	-30,56	27,83	21,65	12,12	12,12	-23,30
18	142,4467391	-28,67	-35,69	-63,35	-64,63	-8,22	-12,29	-23,94	-23,94	-31,44
19	197,8879348	26,77	55,44	-110,88	-7,57	-6,37	-8,33	-22,09	-22,09	2,85
20	65,25945652	-105,85	-132,63	-31,49	-138,58	-6,73	-6,56	-22,44	-22,44	-78,03
21	111,9021739	-59,21	46,64	8,12	-90,32	7,67	9,95	-8,05	-8,05	-0,22
22	149,1652174	-21,95	37,26	-9,20	-51,43	32,56	36,96	16,84	16,84	-0,79
23	144,2648913	-26,85	-4,90	-31,97	-54,72	-7,50	-0,98	-23,22	-23,22	-21,12
24	165,1720652	-5,94	20,91	-26,44	-32,19	21,92	30,56	6,20	6,20	0,12
25	205,3306522	34,22	40,16	-21,76	9,59	34,20	28,18	2,76	2,76	21,79
26	274,1422826	103,03	68,81	99,52	80,02	67,12	61,10	35,69	35,69	92,11
27	288,1923913	117,08	14,05	148,51	95,69	105,29	99,27	73,85	73,85	75,91
28	68,55586957	-102,56	-219,64	-83,32	-122,33	-25,56	-31,58	-56,99	-56,99	-159,21
29	70,06065217	-101,05	1,50	-108,07	-119,20	-71,23	-77,25	-102,66	-102,66	-57,99
30	87,30225	-83,81	17,24	-55,14	-100,34	-54,36	-60,38	-85,79	-85,79	-26,97
31	142,2207609	-28,89	54,92	-55,67	-43,80	-53,03	-59,05	-84,46	-84,46	-2,15
32	69,52163043	-101,59	-72,70	4,26	-114,88	6,55	0,52	-24,89	-24,89	-53,52
33	57,57630435	-113,54	-11,95	-54,33	-125,21	-37,64	-43,67	-69,08	-69,08	-63,59
34	86,00554348	-85,11	28,43	-63,16	-95,16	-21,59	-27,61	-53,02	-53,02	-33,84
35	101,4892391	-69,62	15,48	-42,78	-78,06	-41,27	-47,29	-72,70	-72,70	-25,20
36	83,36923913	-87,74	-18,12	-81,80	-94,56	-50,87	-56,89	-82,31	-82,31	-57,76
37	103,1652174	-67,95	19,80	-102,17	-73,14	-90,39	-81,75	-106,11	-106,11	-38,25
38	193,9621739	22,85	90,80	-80,18	19,28	-35,48	-28,96	-51,20	-51,20	28,17
39	110,6206522	-60,49	-83,34	-177,57	-62,45	-94,71	-90,31	-110,42	-110,42	-93,19
40	129,2561957	-41,86	18,64	60,70	-42,19	12,72	15,01	-2,99	-2,99	33,50
41	156,4667391	-14,65	27,21	86,41	-13,36	-7,24	-7,08	-22,96	-22,96	19,54
42	187,2211957	16,11	30,75	99,92	19,01	23,14	21,19	7,42	7,42	32,23
43	221,6002174	50,49	34,38	79,38	55,01	3,93	-0,14	-11,79	-11,79	37,82
44	127,8379348	-43,28	-93,76	58,32	-37,13	42,44	36,25	26,72	26,72	-42,77
45	203,361087	32,25	75,52	145,78	40,01	85,72	77,41	70,00	70,00	71,01
46	169,6195652	-1,49	-33,74	83,61	7,89	39,60	29,18	23,89	23,89	-9,38
47	219,5959783	48,48	49,98	118,11	59,49	54,42	41,88	38,70	38,70	56,32
48	249,1634783	78,05	29,57	165,79	90,67	92,50	77,84	76,78	76,78	69,48
49	242,6421739	71,53	-6,52	139,48	85,77	-4,77	24,54	26,66	26,66	40,16
50	295,3003261	124,19	52,66	101,34	140,05	12,00	37,08	43,43	43,43	71,59
51	345,5905435	174,48	50,29	234,97	191,96	86,41	107,25	117,84	117,84	134,62
52	191,521087	20,41	-154,07	62,26	39,51	21,13	37,74	52,56	52,56	-61,31
53	201,4856522	30,37	9,96	45,02	51,09	-16,08	-3,71	15,35	15,35	17,42
54	242,695	71,58	41,21	55,47	93,92	24,76	32,90	56,19	56,19	48,54
55	262,9501087	91,84	20,26	41,35	115,80	-8,58	-4,68	22,85	22,85	43,07
56	112,6777174	-58,44	-150,27	-15,16	-32,86	-26,58	-26,91	4,86	4,86	-83,88
57	156,6558696	-14,46	43,98	-46,71	12,74	-14,84	-19,41	16,59	16,59	-3,69
58	131,988587	-39,12	-24,67	-37,63	-10,31	-51,89	-60,69	-20,45	-20,45	-28,51
59	229,3596739	58,25	97,37	9,76	88,68	10,33	-2,71	41,76	41,76	60,88
60	139,056413	-32,06	-90,30	-110,11	0,00	-71,46	-88,74	-40,03	-40,03	-77,99
xbarra		0,00	-1,62	6,05	-15,99	0,00	0,00	0,00	0,00	-1,09
61	209,1434783	38,03	70,09	-33,50	71,71	-123,56	-67,54	-13,54	-13,54	41,84
62	142,7758696	-28,34	3,72	-152,52	6,96	-225,81	-176,15	-115,80	-115,80	-54,87
63	188,9634783	17,85	49,91	-156,63	54,77	-155,51	-112,20	-45,49	-45,49	-30,78
64	80,54163043	-90,57	-58,51	-110,98	-52,04	-175,14	-138,18	-65,12	-65,12	-96,38
65	160,1268478	-10,99	21,07	-41,36	29,17	-142,73	-112,13	-32,71	-32,71	-21,41
66	212,5019565	41,39	73,45	-30,19	83,16	-90,73	-66,48	19,29	19,29	17,82
67	256,3836957	85,27	117,33	-6,57	128,66	-100,44	-82,54	9,58	9,58	55,25
68	101,9643478	-69,15	-37,09	-10,71	-24,14	-122,58	-111,04	-12,56	-12,56	-54,28
69	177,1534783	6,04	38,10	20,50	52,67	-79,64	-74,45	30,38	30,38	7,63
70	188,3555435	17,24	49,30	56,37	65,49	-80,81	-81,97	29,21	29,21	26,19
71	222,7878261	51,67	83,73	-6,57	101,55	-81,54	-89,05	28,48	28,48	31,40
72	239,8611957	68,75	100,80	100,80	120,24	-55,95	-69,82	54,07	54,07	75,55

9.5.3. Total líquido por cliente – *cluster 2*

Período	Observado	Resíduos								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	3122,126667	-1637,83	-	-	-	-507,85	-1125,86	453,68	453,68	3,12
2	3383,83	-1376,13	261,70	-	190,94	67,76	-412,91	1029,29	1029,29	166,54
3	2471,72	-2288,24	-912,11	-	-791,93	-1282,82	-1626,14	-321,29	-321,29	-557,11
4	3049,473333	-1710,48	577,75	-	-284,94	-444,09	-650,05	517,44	517,44	176,25
5	4509,146667	-250,81	1459,67	-	1103,97	-423,51	-492,12	538,02	538,02	1108,28
6	4543,276667	-216,68	34,13	-	1067,34	261,68	330,42	1223,21	1223,21	598,81
7	6265,286667	1505,33	1722,01	-	2718,59	760,37	966,47	1721,90	1721,90	1879,39
8	2386,766667	-2373,19	-3878,52	-	-1230,69	460,96	804,41	1422,49	1422,49	-1499,81
9	3279,766667	-1480,19	893,00	-	-408,45	-186,91	293,88	774,61	774,61	-93,56
10	5607,096667	847,14	2327,33	-	1848,12	614,68	1232,83	1576,21	1576,21	1485,20
11	5222,38	462,42	-384,72	-	1392,64	-781,78	-26,28	179,75	179,75	-19,05
12	3429,16	-1330,80	-1793,22	-	-471,34	-794,80	98,06	166,73	166,73	-1308,73
13	3355,63	-1404,33	-73,53	233,50	-615,63	370,80	-75,62	-109,96	-109,96	-14,71
14	3453,403333	-1306,55	97,77	69,57	-588,62	782,48	404,73	301,71	301,71	-448,42
15	3089,736667	-1670,22	-363,67	618,02	-1023,05	-19,66	-328,73	-500,42	-500,42	-495,80
16	3218,333333	-1541,62	128,60	168,86	-965,21	369,92	129,52	-110,85	-110,85	-424,40
17	5222,73	462,77	2004,40	713,58	968,42	935,22	763,50	454,46	454,46	1006,64
18	4021,233333	-738,72	-1201,50	-522,04	-303,84	384,78	281,74	-95,98	-95,98	-625,36
19	7346,996667	2587,04	3325,76	1081,71	2951,16	2487,23	2452,86	2006,47	2006,47	1948,81
20	1655,776667	-3104,18	-5691,22	-730,99	-2810,82	375,12	409,43	-105,64	-105,64	-2007,17
21	3318,21	-1441,75	1662,43	38,44	-1219,14	496,68	599,66	15,91	15,91	-37,54
22	4377,45	-382,51	1059,24	-1229,65	-230,67	30,18	201,84	-450,58	-450,58	-577,16
23	4357,183333	-402,77	-20,27	-865,20	-321,69	-1001,83	-761,49	-1482,59	-1482,59	-542,05
24	3230,413333	-1529,54	-1126,77	-198,75	-1519,23	-348,39	-39,39	-829,16	-829,16	-336,81
25	4161,486667	-598,47	931,07	805,86	-658,91	860,28	722,86	-101,25	-101,25	1169,28
26	2989,12	-1770,84	-1172,37	-464,28	-1902,04	1,82	-135,60	-959,71	-959,71	-643,94
27	4743,633333	-16,32	1754,51	1653,90	-218,29	1317,86	1180,44	356,33	356,33	1826,86
28	2572,063333	-2187,89	-2171,57	-646,27	-2460,62	-592,73	-730,15	-1554,26	-1554,26	-1186,01
29	3776,796667	-983,16	1204,73	-1445,93	-1326,65	-827,09	-964,51	-1788,62	-1788,62	-808,75
30	3783,336667	-976,62	6,54	-237,90	-1390,87	-169,49	-306,91	-1131,02	-1131,02	305,65
31	5878,16	1118,20	2094,82	-1468,84	633,19	702,01	564,60	-259,51	-259,51	-274,47
32	2475,77	-2284,19	-3402,39	819,99	-2839,96	878,73	741,32	-82,79	-82,79	-4,82
33	1998,64	-2761,32	-477,13	-1319,57	-3387,85	-1139,27	-1276,69	-2100,80	-2100,80	-1039,05
34	3132,34	-1627,62	1133,70	-1245,11	-2324,91	-1531,31	-1668,73	-2492,84	-2492,84	-148,86
35	4693,27	-66,69	1560,93	336,09	-834,74	-982,12	-1119,54	-1943,65	-1943,65	1197,43
36	4326,71	-433,25	-366,56	1096,30	-1272,07	431,52	294,10	-530,01	-530,01	790,17
37	4413,866667	-346,09	87,16	252,38	-1255,67	-165,25	143,76	-646,01	-646,01	295,91
38	3104,11	-1655,85	-1309,76	114,99	-2636,19	-1161,10	-920,77	-1641,87	-1641,87	250,92
39	4619,236667	-140,72	1515,13	-124,40	-1191,82	-84,45	87,21	-565,21	-565,21	401,03
40	3933,613333	-826,34	-685,62	1361,55	-1948,21	-509,09	-406,11	-989,86	-989,86	503,16
41	4833,793333	73,84	900,18	1057,00	-1118,79	-1048,00	-1013,70	-1528,77	-1528,77	313,11
42	4948,133333	188,18	114,34	1164,80	-1075,21	-282,61	-316,98	-763,37	-763,37	586,25
43	4519,613333	-240,34	-428,52	-1358,55	-1574,49	-1934,44	-2037,49	-2415,21	-2415,21	-1192,54
44	1984,676667	-2775,28	-2534,94	-491,09	-4180,19	-890,27	-1061,99	-1371,03	-1371,03	-1074,12
45	3413,453333	-1346,50	1428,78	1414,81	-2822,18	-1002,37	-1242,77	-1483,13	-1483,13	747,00
46	6319,016667	1559,06	2905,56	3186,68	12,63	377,46	68,38	-103,31	-103,31	2143,74
47	9363,473333	4603,52	3044,46	4670,20	2986,32	2410,17	2032,42	1929,41	1929,41	3529,75
48	6000,253333	1240,30	-3363,22	1673,54	-447,66	827,16	380,73	346,39	346,39	-51,65
49	6260,563333	1500,61	260,31	1846,70	-258,11	-557,99	334,86	403,54	403,54	943,79
50	6813,703333	2053,75	553,14	3709,59	224,27	309,05	1064,55	1270,58	1270,58	1588,10
51	7012,193333	2252,24	198,49	2392,96	352,00	69,07	687,22	1030,60	1030,60	-921,88
52	7858,146667	3098,19	845,95	3924,53	1127,19	1176,00	1656,80	2137,53	2137,53	802,75
53	9484,62	4724,66	1626,47	4650,83	2682,90	1363,38	1706,83	2324,91	2324,91	2185,79
54	7275,82	2515,86	-2208,80	2327,69	403,34	-194,36	11,74	767,17	767,17	-1061,14
55	6678,313333	1918,36	-597,51	2158,70	-264,93	-2015,18	-1946,44	-1053,65	-1053,65	-691,69
56	4289,83	-470,13	-2388,48	2305,15	-2724,17	-824,55	-893,16	136,98	136,98	-1189,67
57	8487,143333	3727,19	4197,31	5073,69	1402,38	1831,88	1625,92	2793,41	2793,41	1780,64
58	8689,99	3930,03	202,85	2370,97	1534,46	508,99	165,68	1470,52	1470,52	-892,40
59	9548,283333	4788,33	858,29	184,81	2321,99	355,55	-125,12	1317,08	1317,08	-948,42
60	7297,05	2537,09	-2251,23	1296,80	0,00	-115,49	-733,50	846,04	846,04	-80,90
xbarra		0,00	70,76	883,87	-457,25	0,00	0,00	0,00	0,00	142,27
61	6762,336667	2002,38	-534,71	501,77	-605,47	-3257,18	-1643,07	108,17	108,17	-1294,08
62	5617,143333	857,19	-1679,91	-1196,56	-1821,43	-4088,47	-2680,39	-723,12	-723,12	-3500,60
63	7566,52	2806,56	269,47	554,33	57,19	-2577,57	-1375,51	787,78	787,78	-1305,08
64	6381,823333	1621,87	-915,23	-1476,32	-1198,27	-3501,29	-2505,25	-135,94	-135,94	-2626,58
65	4277,43	-482,53	-3019,62	-5207,19	-3373,43	-7044,77	-6254,77	-3679,42	-3679,42	-5696,48
66	5062,523333	302,57	-2234,53	-2213,30	-2659,10	-5608,62	-5024,64	-2243,27	-2243,27	-3742,17
67	8002,38	3242,42	705,33	1324,07	210,00	-3892,08	-3514,13	-526,73	-526,73	-614,51
68	3575,89	-1184,07	-3721,16	-713,94	-4287,25	-4739,46	-4567,54	-1374,11	-1374,11	-3400,11
69	5576,136667	816,18	-1720,91	-2911,01	-2357,77	-4280,09	-4314,20	-914,74	-914,74	-3964,88
70	7948,49	3188,53	651,44	-741,50	-56,17	-3433,47	-3673,61	-68,12	-68,12	-1689,83
71	6126,62	1366,66	-1170,43	-3421,66	-1948,81	-6267,08	-6713,24	-2901,73	-2901,73	-4159,73
72	4530,43	-229,53	-2766,62	-2766,62	-3615,76	-6083,07	-6735,26	-2717,72	-2717,72	-4363,93

9.5.4. Total líquido por tipo de lavagem – *cluster* 1

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	17,25538462	-71,79	-	-	-	-75,82	-67,71	-90,22	-90,22	-71,79
2	24,54153846	-64,50	7,29	-	7,45	-117,91	-111,75	-132,30	-132,30	-64,50
3	150,3892308	61,34	125,85	-	133,47	34,54	38,74	20,14	20,14	61,34
4	25,14692308	-63,90	-125,24	-	8,39	-18,96	-16,72	-33,36	-33,36	-63,90
5	217,2192308	128,17	192,07	-	200,63	128,82	129,11	114,42	114,42	128,17
6	172,93	83,89	-44,29	-	156,51	98,82	97,14	84,42	84,42	83,89
7	92,41692308	3,37	-80,51	-	76,16	-22,02	-25,66	-36,42	-36,42	3,37
8	189,3176923	100,27	96,90	-	173,23	116,44	110,85	102,04	102,04	100,27
9	13,52	-75,52	-175,80	-	-2,40	-66,59	-74,14	-80,99	-80,99	-75,52
10	62,11153846	-26,93	48,59	-	46,36	9,27	-0,23	-5,13	-5,13	-26,93
11	50,53307692	-38,51	-11,58	-	34,95	-19,57	-31,04	-33,97	-33,97	-38,51
12	38,70923077	-50,34	-11,82	-	23,29	-39,56	-52,98	-53,96	-53,96	-50,34
13	101,1892308	12,14	62,48	83,93	85,94	-8,03	-1,32	-0,83	-0,83	12,14
14	120,4946154	31,45	19,31	95,95	105,41	-38,10	-32,37	-30,90	-30,90	31,45
15	70,83769231	-18,21	-49,66	-79,55	55,92	-61,16	-56,41	-53,96	-53,96	-18,21
16	39,60153846	-49,44	-31,24	14,45	24,85	-20,65	-16,88	-13,45	-13,45	-49,44
17	123,3346154	34,29	83,73	-93,88	108,75	18,79	21,59	25,99	25,99	34,29
18	68,11461538	-20,93	-55,22	-104,82	53,70	-22,14	-20,33	-14,94	-14,94	-20,93
19	195,9638462	106,92	127,85	103,55	181,71	65,38	66,21	72,58	72,58	106,92
20	40,97538462	-48,07	-154,99	-148,34	26,89	-48,05	-48,19	-40,85	-40,85	-48,07
21	17,12	-71,92	-23,86	3,60	3,20	-79,13	-80,25	-71,93	-71,93	-71,92
22	97,56307692	8,52	80,44	35,45	83,81	28,58	26,48	35,78	35,78	8,52
23	69,33384615	-19,71	-28,23	18,80	55,75	-16,92	-20,00	-9,72	-9,72	-19,71
24	221,4761538	132,43	152,14	182,77	208,06	127,06	123,00	134,26	134,26	132,43
25	199,3715385	110,33	-22,10	98,18	186,12	88,40	91,06	102,80	102,80	110,33
26	257,6646154	168,62	58,29	137,17	244,58	97,33	99,98	111,73	111,73	168,62
27	225,0561538	136,01	-32,61	154,22	212,14	91,31	93,97	105,71	105,71	136,01
28	35,08153846	-53,96	-189,97	-4,52	22,33	-26,92	-24,26	-12,52	-12,52	-53,96
29	52,41923077	-36,63	17,34	-70,92	39,84	-53,87	-51,21	-39,47	-39,47	-36,63
30	83,52461538	-5,52	31,11	15,41	71,11	-8,48	-5,82	5,92	5,92	-5,52
31	215,6692308	126,62	132,14	19,71	203,42	83,34	85,99	97,74	97,74	126,62
32	57,75692308	-31,29	-157,91	16,78	45,68	-33,01	-30,36	-18,61	-18,61	-31,29
33	14,14230769	-74,90	-43,61	-2,98	2,23	-83,86	-81,20	-69,46	-69,46	-74,90
34	34,58	-54,46	20,44	-62,98	22,84	-36,15	-33,50	-21,75	-21,75	-54,46
35	45,79538462	-43,25	11,22	-23,54	34,22	-42,20	-39,55	-27,80	-27,80	-43,25
36	18,63769231	-70,41	-27,16	-202,84	7,23	-77,53	-74,87	-63,13	-63,13	-70,41
37	89,11615385	0,07	70,48	-110,26	77,87	-9,20	-13,26	-2,00	-2,00	0,07
38	304,8361538	215,79	215,72	47,17	293,76	157,15	154,07	164,35	164,35	215,79
39	92,54230769	3,50	-212,29	-132,51	81,63	-28,55	-30,65	-21,35	-21,35	3,50
40	88,03769231	-1,01	-4,50	52,96	77,29	38,69	37,57	45,89	45,89	-1,01
41	18,18538462	-70,86	-69,85	-34,23	7,61	-75,45	-75,59	-68,25	-68,25	-70,86
42	41,57307692	-47,47	23,39	-41,95	31,16	-37,78	-36,94	-30,58	-30,58	-47,47
43	27,47461538	-61,57	-14,10	-188,19	17,23	-92,21	-90,39	-85,01	-85,01	-61,57
44	60,11153846	-28,93	32,64	2,35	50,04	-18,01	-15,21	-10,81	-10,81	-28,93
45	129,5376923	40,49	69,43	115,40	119,63	44,19	47,97	51,39	51,39	40,49
46	68,24307692	-20,80	-61,29	33,66	58,50	10,17	14,92	17,37	17,37	-20,80
47	103,8007692	14,76	35,56	58,01	94,23	28,45	34,19	35,65	35,65	14,76
48	122,5923077	33,55	18,79	103,95	113,19	39,08	45,79	46,28	46,28	33,55
49	75,91615385	-13,13	-46,68	-13,20	66,68	4,65	-8,77	-9,75	-9,75	-13,13
50	22,15769231	-66,89	-53,76	-282,68	13,08	-98,47	-109,94	-112,87	-112,87	-66,89
51	57,88923077	-31,16	35,73	-34,65	48,98	-36,15	-45,65	-50,55	-50,55	-31,16
52	50,11846154	-38,93	-7,77	-37,92	41,38	27,83	20,28	13,43	13,43	-38,93
53	48,28076923	-40,76	-1,84	30,10	39,71	-18,30	-23,89	-32,70	-32,70	-40,76
54	21,87384615	-67,17	-26,41	-19,70	13,47	-30,42	-34,06	-44,82	-44,82	-67,17
55	58,14307692	-30,90	36,27	30,67	49,91	-34,48	-36,16	-48,88	-48,88	-30,90
56	33,69615385	-55,35	-24,45	-26,42	25,63	-17,37	-17,09	-31,77	-31,77	-55,35
57	243,6769231	154,63	209,98	114,14	235,77	185,38	187,62	170,99	170,99	154,63
58	19,15153846	-69,89	-224,53	-49,09	11,41	-11,87	-7,67	-26,27	-26,27	-69,89
59	98,53538462	9,49	79,38	-5,27	90,97	50,24	56,40	35,84	35,84	9,49
60	7,403076923	-81,64	-91,13	-115,19	0,00	-49,05	-40,94	-63,45	-63,45	-81,64
xbarra		0,00	-0,17	-6,61	78,02	0,00	0,00	0,00	0,00	0,00
61	19,26153846	-69,78	11,86	-56,65	12,03	-10,55	-35,99	-60,95	-60,95	-69,78
62	22,58846154	-66,46	15,19	0,43	15,52	-56,59	-79,09	-106,99	-106,99	-66,46
63	11,12538462	-77,92	3,72	-46,76	4,22	-41,46	-61,02	-91,86	-91,86	-77,92
64	4,17	-84,87	-3,23	-45,95	-2,57	23,33	6,70	-27,07	-27,07	-84,87
65	21,20769231	-67,84	13,80	-27,07	14,64	-3,92	-17,61	-54,32	-54,32	-67,84
66	123,3346154	34,29	115,93	101,46	116,93	112,49	101,73	62,09	62,09	34,29
67	156,1792308	67,13	148,78	98,04	149,95	105,01	97,18	54,61	54,61	67,13
68	13,64615385	-75,40	6,24	-20,05	7,58	4,04	-0,85	-46,36	-46,36	-75,40
69	0,073846154	-88,97	-7,33	-243,60	-5,83	-16,76	-18,71	-67,16	-67,16	-88,97
70	64,66461538	-24,38	57,26	45,51	58,93	75,10	76,08	24,70	24,70	-24,38
71	110,1738462	21,13	102,77	11,64	104,61	103,33	107,26	52,94	52,94	21,13
72	17,92461538	-71,12	10,52	10,52	12,53	2,92	9,78	-47,48	-47,48	-71,12

9.5.5. Total líquido por tipo de lavagem – *cluster 2*

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	7133,33	-3089,17	-	-	-	1130,05	-52,81	3381,46	3381,46	7,13
2	4952,84	-5269,66	-2180,49	-	-2391,88	-248,30	-1132,53	2003,11	2003,11	-1589,80
3	4315,61	-5906,89	-637,23	-	-3240,49	-590,26	-1175,86	1661,15	1661,15	-1336,83
4	5434,31	-4788,19	1118,70	-	-2333,18	2929,07	2642,11	5180,49	5180,49	152,01
5	6913,53	-3308,97	1479,22	-	-1065,35	1006,26	1017,93	3257,68	3257,68	1368,19
6	7885,54	-2336,96	972,01	-	-304,73	118,48	428,79	2369,90	2369,90	1714,82
7	8464,27	-1758,23	578,73	-	62,62	598,51	1207,44	2849,92	2849,92	1605,40
8	1013,44	-9209,06	-7450,83	-	-7599,60	299,38	1206,95	2550,80	2550,80	-5429,71
9	2885,38	-7337,12	1871,94	-	-5939,05	-347,98	858,22	1903,43	1903,43	-1884,70
10	4783,43	-5439,07	1898,05	-	-4252,39	-430,32	1074,51	1821,09	1821,09	429,13
11	7631,37	-2591,13	2847,94	-	-1615,83	-2504,53	-701,07	-253,12	-253,12	2750,67
12	5944,56	-4277,94	-1686,81	-	-3514,03	-2346,60	-244,50	-95,18	-95,18	314,81
13	3878,89	-6343,61	-2065,67	-3254,44	-5791,09	-2161,57	-3212,62	-3287,27	-3287,27	-2228,22
14	3932,41	-6290,09	53,52	-1020,43	-5948,96	-1305,91	-2207,64	-2431,61	-2431,61	-183,51
15	5148,14	-5074,36	1215,73	832,53	-4944,61	205,09	-547,33	-920,62	-920,62	1398,79
16	4036,6	-6185,90	-1111,54	-1397,71	-6267,54	1494,19	891,09	368,48	368,48	-768,60
17	6477,2	-3745,30	2440,60	-436,33	-4038,33	532,76	78,97	-592,95	-592,95	1209,79
18	7285,62	-2936,88	808,42	-599,92	-3441,30	-518,61	-823,08	-1644,32	-1644,32	1106,51
19	7097,86	-3124,64	-187,76	-1366,41	-3840,44	-805,08	-960,23	-1930,78	-1930,78	240,47
20	1047,58	-9174,92	-6050,28	34,14	-10102,11	296,35	290,51	-829,36	-829,36	-2199,91
21	3880,55	-6341,95	2832,97	995,17	-7480,53	610,01	753,49	-515,69	-515,69	481,57
22	9492,57	-729,93	5612,02	4709,14	-2079,90	4241,64	4534,44	3115,93	3115,93	4982,30
23	10018,24	-204,26	525,67	2386,87	-1765,61	-154,84	287,28	-1280,54	-1280,54	2337,06
24	8686,8	-1535,70	-1331,44	2742,24	-3308,44	358,47	949,90	-767,24	-767,24	1017,07
25	8345,28	-1877,22	-341,52	4466,39	-3861,35	16,23	-443,39	-2235,18	-2235,18	1340,80
26	5189,84	-5032,66	-3155,44	1257,43	-7228,18	-2337,07	-2796,69	-4588,48	-4588,48	-2314,45
27	10297,29	74,79	5107,45	5149,15	-2332,11	3065,65	2606,03	814,24	814,24	3007,13
28	3260,07	-6962,43	-7037,22	-776,53	-9580,72	-1570,93	-2030,55	-3822,35	-3822,35	-4540,39
29	5958,41	-4264,09	2698,34	-518,79	-7093,77	-2274,62	-2734,24	-4526,04	-4526,04	-1449,55
30	7283,36	-2939,14	1324,95	-2,26	-5980,21	-2809,46	-3269,08	-5060,88	-5060,88	-13,79
31	10382,66	160,16	3099,30	3284,80	-3092,29	191,13	-268,49	-2060,28	-2060,28	3183,48
32	5248,57	-4973,93	-5134,09	4200,99	-8437,77	2208,75	1749,13	-42,67	-42,67	-74,29
33	5512,27	-4710,23	263,70	1631,72	-8385,46	-46,86	-506,48	-2298,27	-2298,27	-1189,09
34	5310,3	-4912,20	-201,97	-4182,27	-8798,82	-2229,22	-2688,84	-4480,63	-4480,63	-3754,05
35	11377,35	1154,85	6067,05	1359,11	-2943,15	-1084,32	-1543,94	-3335,73	-3335,73	3377,45
36	11427,54	1205,04	50,19	2740,74	-3104,35	810,62	351,00	-1440,80	-1440,80	2895,66
37	11883,21	1660,71	455,67	3537,93	-2860,07	-985,84	-394,41	-2111,55	-2111,55	2498,56
38	11597,86	1375,36	-285,35	6408,02	-3356,81	-469,05	-26,94	-1594,76	-1594,76	2895,89
39	5653,92	-4568,58	-5943,94	-4643,37	-9512,13	-6117,72	-5824,93	-7243,43	-7243,43	-6602,76
40	7390,29	-2832,21	1736,37	4130,22	-7987,15	-1980,72	-1837,23	-3106,42	-3106,42	953,65
41	13966,7	3744,20	6576,41	8008,29	-1622,13	1193,66	1187,83	67,96	67,96	5855,47
42	19100,49	8877,99	5133,79	11817,13	3300,27	4467,66	4312,51	3341,95	3341,95	8265,42
43	14352,56	4130,06	-4747,93	3969,90	-1659,04	-378,97	-683,44	-1504,68	-1504,68	-936,81
44	9883,37	-339,13	-4469,19	4634,80	-6339,62	2303,54	1849,76	1177,83	1177,83	-2529,60
45	9871,69	-350,81	-11,68	4359,42	-6562,69	-227,44	-830,54	-1353,15	-1353,15	-1779,14
46	11203,71	981,21	1332,02	5893,41	-5442,06	-875,81	-1628,23	-2001,52	-2001,52	281,03
47	21839,35	11616,85	10635,64	10462,00	4982,20	4837,68	3935,95	3711,97	3711,97	7809,13
48	18677,59	8455,09	-3161,76	7250,05	1609,05	3520,66	2469,61	2394,96	2394,96	1866,02
49	21661,6	11439,10	2984,01	9778,39	4381,67	2001,13	4103,23	4252,54	4252,54	3965,42
50	23218,66	12996,16	1557,06	11620,80	5727,34	4360,33	6163,79	6611,74	6611,74	4284,26
51	22000,31	11777,81	-1218,35	16346,39	4297,61	3437,25	4942,08	5688,66	5688,66	4468,01
52	15290,81	5068,31	-6709,50	7900,52	-2623,28	-871,61	334,59	1379,80	1379,80	-4666,64
53	19106,39	8883,89	3815,58	5139,69	980,91	-458,06	449,50	1793,35	1793,35	-2464,74
54	20166,18	9943,68	1059,79	1065,69	1829,31	-1258,07	-649,13	993,35	993,35	-3080,53
55	21917,36	11694,86	1751,18	7564,80	3369,11	394,41	704,71	2645,82	2645,82	2112,27
56	9263,23	-959,27	-12654,13	-620,14	-9496,41	-5108,02	-5096,35	-2856,60	-2856,60	-9071,41
57	16902,82	6680,32	7639,59	7031,13	-2068,21	12,27	-274,69	2263,68	2263,68	1776,00
58	18164,66	7942,16	1261,84	6960,95	-1017,76	-706,28	-1291,88	1545,13	1545,13	1750,46
59	22699,1	12476,60	4534,44	859,75	3305,30	-1093,99	-1978,22	1157,42	1157,42	393,36
60	19605,19	9382,69	-3093,91	927,60	0,00	-2343,15	-3526,02	-91,74	-91,74	-12711,65
xbarra		0,00	211,39	3388,31	-3200,09	0,00	0,00	0,00	0,00	378,92
61	20001,01	9778,51	395,82	-1660,59	184,43	-8702,29	-4629,91	-822,34	-822,34	-1906,58
62	12472,96	2250,46	-7132,23	-10745,70	-7555,00	-15428,20	-11803,77	-7548,25	-7548,25	-10206,70
63	13118,16	2895,66	-6487,03	-8882,15	-7121,19	-14487,73	-11311,25	-6607,79	-6607,79	-8957,38
64	7594,1	-2628,40	-12011,09	-7696,71	-12856,64	-17611,16	-14882,62	-9731,21	-9731,21	-11154,54
65	9239,87	-982,63	-10365,32	-9866,52	-11422,26	-19367,42	-17086,83	-11487,47	-11487,47	-11400,72
66	5882,87	-4339,63	-13722,32	-14283,31	-14990,64	-24584,21	-22751,57	-16704,26	-16704,26	-15283,22
67	12808,68	2586,18	-6796,51	-9108,68	-8276,22	-17757,10	-16372,41	-9877,15	-9877,15	-9225,73
68	5360,91	-4861,59	-14244,28	-3902,32	-15935,38	-18053,17	-17116,43	-10173,22	-10173,22	-10398,96
69	8813,92	-1408,58	-10791,27	-8088,90	-12693,76	-17119,46	-16630,67	-9239,51	-9239,51	-10734,03
70	13036,71	2814,21	-6568,48	-5127,95	-8682,35	-14877,06	-14836,22	-6997,12	-6997,12	-17136,92
71	11714,28	1491,78	-7890,91	-10984,82	-10216,17	-21121,64	-21528,75	-13241,69	-13241,69	-10707,75
72	14714,92	4492,42	-4890,27	-4890,27	-7426,92	-16276,26	-17131,31	-8396,31	-8396,31	-6173,00

9.5.6. Total líquido por tipo de lavagem – *cluster 3*

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	2923,26	-3308,81	-	-	-	-1736,82	-1904,76	-638,50	-638,50	2,92
2	3275,42	-2956,65	352,16	-	286,75	-1415,11	-1472,94	-316,79	-316,79	273,73
3	2515,01	-3717,06	-760,41	-	-539,07	-2262,10	-2209,81	-1163,78	-1163,78	-524,06
4	2485,76	-3746,31	-29,25	-	-633,74	-2666,97	-2504,58	-1568,65	-1568,65	-320,46
5	3649,04	-2583,03	1163,28	-	464,13	-2597,73	-2325,23	-1499,42	-1499,42	859,97
6	3418,36	-2813,71	-230,68	-	168,04	-2536,09	-2153,48	-1437,77	-1437,77	300,73
7	8371,98	2139,91	4953,62	-	5056,25	479,46	972,18	1577,78	1577,78	4658,53
8	3606,51	-2625,56	-4765,47	-	225,36	821,25	1424,07	1919,56	1919,56	-1536,54
9	6748,22	516,15	3141,71	-	3301,66	-98,71	614,23	999,61	999,61	1927,00
10	9746,53	3514,46	2998,31	-	6234,56	2672,03	3495,08	3770,35	3770,35	3861,82
11	7477,34	1245,27	-2269,19	-	3899,96	558,96	1492,11	1657,28	1657,28	248,79
12	5895,15	-336,92	-1582,19	-	2252,36	1784,88	2828,14	2883,19	2883,19	-1283,42
13	5366,2	-865,87	-528,95	2442,94	1657,99	1484,63	963,00	935,47	935,47	-1198,96
14	5733,21	-498,86	367,01	2457,79	1959,59	1821,20	1354,62	1272,04	1272,04	-442,33
15	3388,4	-2843,67	-2344,81	873,39	-450,63	-610,19	-1021,71	-1159,35	-1159,35	-2262,40
16	5179,64	-1052,43	1791,24	2693,88	1275,20	805,43	448,96	256,27	256,27	463,69
17	8640,95	2408,88	3461,31	4991,91	4671,09	3172,69	2871,28	2623,53	2623,53	3238,94
18	7560,29	1328,22	-1080,66	4141,93	3525,02	2384,36	2138,00	1835,20	1835,20	946,37
19	13021,9	6789,83	5461,61	4649,92	8921,22	5907,90	5716,59	5358,74	5358,74	3916,94
20	2661,36	-3570,71	-10360,54	-945,15	-1504,73	654,61	518,36	105,45	105,45	-5997,28
21	8063,75	1831,68	5402,39	1315,53	3832,25	1995,34	1914,14	1446,18	1446,18	490,08
22	7485,48	1253,41	-578,27	-2261,05	3188,56	1189,50	1163,35	640,34	640,34	-1562,15
23	4345,66	-1886,41	-3139,82	-3131,68	-16,67	-1794,21	-1765,29	-2343,37	-2343,37	-3105,36
24	3369,86	-2862,21	-975,80	-2525,29	-1057,88	38,10	122,07	-511,06	-511,06	-2153,74
25	3983,06	-2249,01	613,20	-1383,14	-510,09	-218,31	-655,97	-1316,63	-1316,63	-445,67
26	4094,58	-2137,49	111,52	-1638,63	-463,99	-137,24	-574,90	-1235,55	-1235,55	-310,37
27	4809,61	-1422,46	715,03	1421,21	185,63	491,22	53,55	-607,10	-607,10	1527,93
28	4468,62	-1763,45	-340,99	-711,02	-220,77	-225,39	-663,06	-1323,71	-1323,71	-192,43
29	5301,01	-931,06	832,39	-3339,94	546,21	-487,05	-924,72	-1585,37	-1585,37	-755,43
30	6293,48	61,41	992,47	-1266,81	1473,27	797,75	360,08	-300,57	-300,57	1002,03
31	8606,47	2374,40	2312,99	-4415,43	3720,84	1172,66	735,00	74,35	74,35	585,95
32	3726,41	-2505,66	-4880,06	1065,05	-1224,63	1399,86	962,19	301,54	301,54	-121,38
33	2340,82	-3891,25	-1385,59	-5722,93	-2675,63	-4047,40	-4485,06	-5145,71	-5145,71	-3757,16
34	3767,69	-2464,38	1426,87	-3717,79	-1314,17	-2848,10	-3285,76	-3946,41	-3946,41	-566,92
35	3303,4	-2928,67	-464,29	-1042,26	-1843,87	-3156,27	-3593,93	-4254,59	-4254,59	534,01
36	1743,49	-4488,58	-1559,91	-1626,37	-3469,20	-1908,07	-2345,74	-3006,39	-3006,39	-826,26
37	4957,94	-1274,13	3214,45	974,88	-320,16	-661,55	-577,58	-1210,71	-1210,71	2460,89
38	4199,35	-2032,72	-758,59	104,77	-1144,16	-1450,59	-1421,67	-1999,74	-1999,74	661,05
39	3784,61	-2447,46	-414,74	-1025,00	-1624,31	-1951,90	-1978,04	-2501,06	-2501,06	-324,96
40	7749,76	1517,69	3965,15	3281,14	2275,42	1637,63	1556,43	1088,47	1088,47	3916,54
41	4791,07	-1441,00	-2958,69	-509,94	-748,68	-2415,11	-2551,36	-2964,27	-2964,27	-978,04
42	5079,66	-1152,41	288,59	-1213,82	-525,50	-1834,19	-2025,50	-2383,35	-2383,35	-716,87
43	5635,78	-596,29	556,12	-2970,69	-34,79	-3216,15	-3462,51	-3765,30	-3765,30	-855,64
44	2019,01	-4213,06	-3616,77	-1707,40	-3716,97	-1725,66	-2027,07	-2274,82	-2274,82	-2049,96
45	9150,09	2918,02	7131,08	6809,27	3348,69	1343,75	987,29	794,60	794,60	6498,66
46	9388,81	3156,74	238,72	5621,12	3522,00	1354,90	943,38	805,75	805,75	3505,73
47	11080,46	4848,39	1691,65	7777,06	5148,24	3202,67	2736,09	2653,51	2653,51	3979,25
48	5958,31	-273,76	-5122,15	4214,82	-39,32	888,63	367,00	339,47	339,47	-2085,98
49	9267,97	3035,90	3309,66	4310,03	3204,92	1132,05	2175,31	2230,36	2230,36	698,03
50	9348,11	3116,04	80,14	5148,76	3219,65	1181,74	2114,89	2280,05	2280,05	819,13
51	12585,92	6353,85	3237,81	8801,31	6392,05	4332,97	5156,02	5431,29	5431,29	3902,66
52	9077,88	2845,81	-3508,04	1328,12	2818,60	449,31	1162,25	1547,63	1547,63	-2868,94
53	12049,82	5817,75	2971,94	7258,75	5725,13	2327,20	2930,03	3425,52	3425,52	2520,77
54	10618,46	4386,39	-1431,36	5538,80	4228,35	1188,17	1680,89	2286,49	2286,49	-51,39
55	7024,49	792,42	-3593,97	1388,71	568,97	-4343,87	-3961,26	-3245,56	-3245,56	-3861,28
56	5111,05	-1121,02	-1913,44	3092,04	-1409,88	-1150,06	-877,56	-51,74	-51,74	-2676,14
57	11129,79	4897,72	6018,74	1979,70	4543,45	807,02	969,41	1905,33	1905,33	1388,39
58	8182,01	1949,94	-2947,78	-1206,80	1530,25	-2368,33	-2316,05	-1270,02	-1270,02	-2221,66
59	11583,07	5351,00	3401,06	502,61	4865,90	1188,85	1131,02	2287,16	2287,16	1356,63
60	6782,58	550,51	-4800,49	824,27	0,00	-803,54	-971,47	294,78	294,78	-1811,77
xbarra		0,00	65,41	1096,85	1402,52	0,00	0,00	0,00	0,00	144,70
61	8295,34	2063,27	1512,76	-972,63	1447,35	-3455,34	-1015,12	388,77	388,77	-975,92
62	11736,82	5504,75	4954,24	2388,71	4823,42	-44,31	2230,75	3799,80	3799,80	2431,45
63	15575,89	9343,82	8793,31	2989,97	8597,07	3708,19	5818,08	7552,30	7552,30	4892,54
64	11641,7	5409,63	4859,12	2563,82	4597,47	-601,62	1343,11	3242,49	3242,49	2451,34
65	9716,53	3484,46	2933,95	-2333,29	2606,89	-3620,84	-1841,27	223,27	223,27	-738,66
66	15429,26	9197,19	8646,68	4810,80	8254,21	2384,22	3998,62	6228,33	6228,33	5583,24
67	17014,45	10782,38	10231,87	9989,96	9773,98	2031,33	3480,58	5875,44	5875,44	8698,00
68	9196,71	2964,64	2414,13	4085,66	1890,83	-679,15	604,93	3164,96	3164,96	1694,60
69	9978,3	3746,23	3195,72	-1151,49	2607,01	-3959,23	-2840,31	-115,12	-115,12	-85,33
70	13113,28	6881,21	6330,70	4931,27	5676,58	-1051,82	-98,06	2792,29	2792,29	4304,20
71	13062,71	6830,64	6280,13	1479,64	5560,60	-946,27	-157,68	2897,84	2897,84	2806,17
72	6785,19	553,12	2,61	2,61	-782,34	-4415,68	-3792,25	-571,57	-571,57	-1428,31

9.5.7. Total líquido por tipo de lavagem – *cluster 4*

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	4505,8	-300,70	-	-	-	759,33	315,79	2127,10	2127,10	4,51
2	6123,61	1317,11	1617,81	-	1624,41	766,22	480,19	2134,00	2134,00	1230,26
3	4745,64	-60,86	-1377,97	-	253,03	-1643,68	-1772,20	-275,91	-275,91	-602,33
4	3431,63	-1374,87	-1314,01	-	-1054,38	-1563,99	-1535,01	-196,22	-196,22	-1409,98
5	4563,22	-243,28	1131,59	-	83,80	-421,24	-234,76	946,53	946,53	252,96
6	4580,18	-226,32	16,96	-	107,36	631,80	975,79	1999,58	1999,58	142,73
7	6565,48	1758,98	1985,30	-	2099,25	804,17	1305,66	2171,94	2171,94	1765,75
8	812,69	-3993,81	-5752,79	-	-3646,94	-1439,67	-780,67	-71,90	-71,90	-4009,83
9	2798,38	-2008,12	1985,69	-	-1654,65	-526,47	290,04	841,30	841,30	-343,72
10	5130,86	324,36	2332,48	-	684,42	-615,38	358,63	752,39	752,39	1814,93
11	8438,32	3631,82	3307,46	-	3998,48	2171,77	3303,29	3539,55	3539,55	3743,48
12	6956,94	2150,44	-1481,38	-	2523,69	1362,79	2651,81	2730,56	2730,56	638,61
13	3081,6	-1724,90	-3875,34	-1424,20	-1345,05	731,61	87,10	47,72	47,72	-2921,08
14	4230,05	-576,45	1148,45	-1893,56	-190,01	269,15	-296,61	-414,74	-414,74	-982,06
15	3820,53	-985,97	-409,52	-925,11	-592,93	-1172,30	-1659,31	-1856,19	-1856,19	-174,73
16	3515,88	-1290,62	-304,65	84,25	-890,98	-83,26	-491,51	-767,14	-767,14	292,46
17	3177,39	-1629,11	-338,49	-1385,83	-1222,88	-410,59	-740,09	-1094,47	-1094,47	-780,39
18	2337,29	-2469,21	-840,10	-2242,89	-2056,38	-214,60	-465,35	-898,49	-898,49	-1245,18
19	7671,91	2865,41	5334,62	1106,43	3284,83	3307,08	3135,09	2623,20	2623,20	3665,69
20	1883,21	-2923,29	-5788,70	1070,52	-2497,27	1027,34	934,09	343,45	343,45	-923,56
21	2444,98	-2361,52	561,77	-353,40	-1928,90	516,62	502,13	-167,27	-167,27	-943,90
22	2734,41	-2072,09	289,43	-2396,45	-1632,88	-1615,35	-1551,08	-2299,23	-2299,23	-1407,57
23	3884,75	-921,75	1150,34	-4553,57	-475,94	-985,31	-842,30	-1669,20	-1669,20	-1291,13
24	3463,57	-1342,93	-421,18	-3493,37	-890,53	-734,09	-512,33	-1417,98	-1417,98	-304,10
25	3378,87	-1427,63	-84,70	297,27	-968,63	1057,59	634,85	-310,18	-310,18	1783,39
26	3030,4	-1776,10	-348,47	-1199,65	-1310,51	-901,79	-1324,53	-2269,56	-2269,56	-41,65
27	4354,66	-451,84	1324,26	534,13	20,35	-609,46	-1032,20	-1977,23	-1977,23	1516,83
28	2487,84	-2318,66	-1866,82	-1028,04	-1839,87	-1082,58	-1505,33	-2450,36	-2450,36	-937,05
29	1820,34	-2986,16	-667,50	-1357,05	-2500,78	-1738,92	-2161,67	-3106,70	-3106,70	-966,99
30	1433,37	-3373,13	-386,97	-903,92	-2881,15	-1089,81	-1512,55	-2457,58	-2457,58	-439,87
31	2151,6	-2654,90	718,23	-5520,31	-2156,33	-2184,52	-2607,26	-3552,29	-3552,29	-2289,53
32	1998,2	-2808,30	-153,40	114,99	-2303,13	1171,04	748,30	-196,73	-196,73	1704,43
33	2420,59	-2385,91	422,39	-24,39	-1874,15	520,94	98,20	-846,83	-846,83	995,28
34	5069,48	262,98	2648,89	2335,07	781,34	748,43	325,69	-619,34	-619,34	3003,58
35	4548,38	-258,12	-521,10	663,63	266,84	-292,97	-715,71	-1660,74	-1660,74	404,76
36	4948,15	141,65	399,77	1484,58	673,20	779,20	356,46	-588,57	-588,57	825,27
37	2488,28	-2318,22	-2459,87	-890,59	-1780,07	-1172,06	-950,29	-1855,94	-1855,94	-1996,31
38	2922,79	-1883,71	434,51	-107,61	-1338,97	-2348,46	-2205,44	-3032,35	-3032,35	-398,03
39	7446,38	2639,88	4523,59	3091,72	3191,22	1143,20	1207,46	459,31	459,31	3630,18
40	4015,35	-791,15	-3431,03	1527,51	-233,22	-894,13	-908,62	-1578,02	-1578,02	-612,05
41	5104,69	298,19	1089,34	3284,35	862,72	206,37	113,12	-477,52	-477,52	1126,71
42	4872,88	66,38	-231,81	3439,51	637,51	1010,64	838,65	326,75	326,75	542,69
43	4324,67	-481,83	-548,21	2173,07	95,89	-1350,51	-1601,25	-2034,39	-2034,39	-647,31
44	2751,12	-2055,38	-1573,55	752,92	-1471,06	584,90	255,40	-98,99	-98,99	-1822,41
45	4095,98	-710,52	1344,86	1675,39	-119,61	857,27	449,02	173,38	173,38	198,36
46	8027,79	3221,29	3931,81	2958,31	3818,80	2367,68	1880,68	1683,80	1683,80	2650,10
47	7279,39	2472,89	-748,40	2731,01	3077,00	1098,98	533,22	415,09	415,09	870,20
48	8198,31	3391,81	918,92	3250,16	4002,51	2690,30	2045,79	2006,41	2006,41	1152,46
49	4990,7	184,20	-3207,61	2502,42	801,50	-1376,47	-87,45	-8,70	-8,70	-1338,14
50	10192,96	5386,46	5202,26	7270,17	6010,35	2214,88	3346,39	3582,65	3582,65	4295,51
51	11292,26	6485,76	1099,30	3845,88	7116,25	2282,25	3256,25	3650,02	3650,02	921,08
52	11240,28	6433,78	-51,98	7224,93	7070,86	3623,96	4440,47	4991,74	4991,74	2206,78
53	9969,54	5163,04	-1270,74	4864,85	5806,72	2364,38	3023,38	3732,16	3732,16	-717,93
54	6231,04	1424,54	-3738,50	1358,16	2074,82	-338,03	163,46	1029,74	1029,74	-3982,32
55	7805,78	2999,28	1574,74	3481,11	3656,15	-576,23	-232,24	791,54	791,54	-162,72
56	3529,45	-1277,05	-4276,33	778,33	-613,58	-1343,60	-1157,12	24,17	24,17	-3537,75
57	4577,19	-229,31	1047,74	481,21	440,75	-1368,35	-1339,38	-0,58	-0,58	-1451,85
58	7481,55	2675,05	2904,36	-546,24	3351,71	-885,39	-1013,92	482,38	482,38	114,53
59	6894,77	2088,27	-586,78	-384,62	2771,52	-1992,47	-2278,50	-624,70	-624,70	-137,98
60	4116,65	-689,85	-2778,12	-4081,66	0,00	-4098,19	-4541,73	-2730,42	-2730,42	-3327,84
xbarra		0,00	-6,60	618,11	503,67	0,00	0,00	0,00	0,00	-10,90
61	3430,47	-1376,03	-686,18	-1560,23	-679,58	-7011,31	-4232,29	-2224,10	-2224,10	-703,08
62	3058,51	-1747,99	-1058,14	-7134,45	-1044,95	-8994,18	-6451,42	-4206,97	-4206,97	-3790,20
63	3747,11	-1059,39	-369,54	-7545,15	-349,75	-9337,51	-7031,01	-4550,31	-4550,31	-3675,35
64	4383,89	-422,61	267,24	-6856,39	293,62	-7307,03	-5236,79	-2519,83	-2519,83	-3011,44
65	1485,49	-3321,01	-2631,16	-8484,05	-2598,18	-10194,27	-8360,29	-5407,07	-5407,07	-5246,61
66	3648,33	-1158,17	-468,32	-2582,71	-428,75	-6995,35	-5397,62	-2208,14	-2208,14	-1132,57
67	2980,25	-1826,25	-1136,40	-4825,53	-1090,23	-9476,36	-8114,90	-4689,16	-4689,16	-2622,54
68	2101,75	-2704,75	-2014,90	-1427,70	-1962,13	-6845,91	-5720,70	-2058,71	-2058,71	-1269,14
69	9310,35	4503,85	5193,70	4733,16	5253,06	-709,80	179,15	4077,40	4077,40	5392,62
70	7994,87	3188,37	3878,22	513,32	3944,18	-4446,67	-3793,98	340,53	340,53	2561,30
71	5730,45	923,95	1613,80	-1164,32	1686,35	-7231,04	-6814,96	-2444,19	-2444,19	603,13
72	4982,84	176,34	866,19	866,19	945,34	-7306,61	-7126,43	-2519,41	-2519,41	1305,48

9.5.8. Total líquido por tipo de lavagem – *cluster 5*

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	1277,83	-1311,98	-	-	-	-1361,52	-1535,79	-1339,21	-1339,21	-1259,80
2	3066,88	477,07	1789,05	-	1793,52	-686,57	-843,75	-664,26	-664,26	800,78
3	1361,12	-1228,69	-1705,76	-	92,24	-2846,52	-2986,60	-2824,21	-2824,21	-1330,94
4	1052,26	-1537,55	-308,86	-	-212,15	-1031,07	-1154,06	-1008,76	-1008,76	-1233,67
5	3076,92	487,11	2024,66	-	1816,99	-959,48	-1065,37	-937,17	-937,17	864,53
6	2505,24	-84,57	-571,68	-	1249,78	-239,93	-328,73	-217,62	-217,62	-189,21
7	6269,7	3679,89	3764,46	-	5018,72	1177,90	1106,19	1200,21	1200,21	3711,37
8	2484,61	-105,20	-3785,09	-	1238,10	-219,62	-274,24	-197,32	-197,32	-970,00
9	3278,51	688,70	793,90	-	2036,47	515,98	478,46	538,28	538,28	725,09
10	4694,96	2105,15	1416,45	-	3457,40	1828,12	1807,70	1850,43	1850,43	1952,52
11	1614,82	-974,99	-3080,14	-	381,73	-1337,47	-1340,80	-1315,16	-1315,16	-1464,86
12	3468,35	878,54	1853,53	-	2239,74	1197,68	1211,44	1219,99	1219,99	1122,02
13	5244,77	2654,96	1776,42	3966,94	4020,63	2920,89	2914,01	2909,74	2909,74	2457,13
14	4711,14	2121,33	-533,63	1644,26	3491,48	1273,16	1274,83	1262,01	1262,01	1500,55
15	4175,6	1585,79	-535,54	2814,48	2960,41	283,44	293,65	272,28	272,28	1092,06
16	3240,92	651,11	-934,68	2188,66	2030,20	1473,07	1491,83	1461,91	1461,91	284,89
17	7492,73	4902,92	4251,81	4415,81	6286,49	3771,81	3799,11	3760,65	3760,65	4759,24
18	2815,08	225,27	-4677,65	309,84	1613,31	385,39	421,24	374,23	374,23	-930,73
19	3655,73	1065,92	840,65	-2613,97	2458,44	-1120,60	-1076,20	-1131,75	-1131,75	1023,63
20	3203,11	613,30	-452,62	718,50	2010,29	814,35	867,30	803,19	803,19	370,86
21	2632,97	43,16	-570,14	-645,54	1444,62	185,91	247,40	174,75	174,75	-91,52
22	2930	340,19	297,03	-1764,96	1746,13	378,63	448,68	367,48	367,48	341,25
23	1375,91	-1213,90	-1554,09	-238,91	196,51	-1260,90	-1182,31	-1272,06	-1272,06	-1283,56
24	1784,97	-804,84	409,06	-1683,38	610,05	-170,23	-83,09	-181,38	-181,38	-504,48
25	2371,75	-218,06	586,78	-2873,02	1201,30	341,04	421,29	318,73	318,73	-15,10
26	6180,73	3590,92	3808,98	1469,59	5014,76	3035,92	3116,17	3013,61	3013,61	3654,18
27	9206,95	6617,14	3026,22	5031,35	8045,45	5607,95	5688,21	5585,64	5585,64	5773,52
28	1202,2	-1387,61	-8004,75	-2038,72	45,17	-272,49	-192,23	-294,80	-294,80	-2951,74
29	1626,78	-963,03	424,58	-5865,95	474,23	-1800,98	-1720,72	-1823,29	-1823,29	-621,31
30	1728,17	-861,64	101,39	-1086,91	580,09	-408,36	-328,10	-430,67	-430,67	-621,01
31	2062,21	-527,60	334,04	-1593,52	918,61	-2420,95	-2340,70	-2443,26	-2443,26	-311,11
32	733,2	-1856,61	-1329,01	-2469,91	-405,93	-1362,40	-1282,14	-1384,71	-1384,71	-1719,65
33	200,3	-2389,51	-532,90	-2432,67	-934,35	-1953,60	-1873,34	-1975,91	-1975,91	-1936,13
34	904,15	-1685,66	703,85	-2025,85	-226,03	-1354,05	-1273,80	-1376,36	-1376,36	-1105,40
35	1264,19	-1325,62	360,04	-111,72	138,48	-1079,46	-999,20	-1101,77	-1101,77	-912,94
36	299,63	-2290,18	-964,56	-1485,34	-821,60	-1362,40	-1282,15	-1384,71	-1384,71	-1963,22
37	218	-2371,81	-81,63	-2153,75	-898,76	-1541,86	-1454,72	-1553,01	-1553,01	-1815,20
38	790,51	-1799,30	572,51	-5390,22	-321,77	-2083,45	-2004,86	-2094,60	-2094,60	-1223,26
39	1490	-1099,81	699,49	-7716,95	382,19	-1838,14	-1768,10	-1849,30	-1849,30	-660,07
40	528,63	-2061,18	-961,37	-673,57	-574,70	-675,20	-613,71	-686,36	-686,36	-1787,99
41	885,58	-1704,23	356,95	-741,20	-213,28	-2271,32	-2218,38	-2282,48	-2282,48	-1202,14
42	786,54	-1803,27	-99,04	-941,63	-307,85	-1079,13	-1034,73	-1090,29	-1090,29	-1386,17
43	7327,65	4737,84	6541,11	5265,44	6237,74	3115,34	3151,19	3104,19	3104,19	5178,52
44	1442,5	-1147,31	-5885,15	709,30	357,06	-382,24	-354,94	-393,40	-393,40	-2264,00
45	3420,96	831,15	1978,46	3220,66	2340,00	1537,92	1556,68	1526,76	1526,76	1115,65
46	1056,13	-1533,68	-2364,83	151,98	-20,36	-931,22	-921,00	-942,37	-942,37	-1720,23
47	2863,17	273,36	1807,04	1598,98	1791,15	790,38	792,04	779,22	779,22	649,85
48	1854,9	-734,91	-1008,27	1555,27	787,36	463,72	456,84	452,57	452,57	-788,65
49	1152,76	-1437,05	-702,14	934,76	89,69	-358,55	-344,79	-336,24	-336,24	-1250,74
50	1086,36	-1503,45	-66,40	295,85	27,77	-1539,05	-1542,39	-1516,74	-1516,74	-1149,96
51	1872,87	-716,94	786,51	382,87	818,75	-1206,73	-1227,15	-1184,42	-1184,42	-347,64
52	1460,98	-1128,83	-411,89	932,35	411,34	505,69	468,17	528,00	528,00	-946,79
53	4168,33	1578,52	2707,35	3282,75	3123,16	1259,97	1205,36	1282,28	1282,28	1858,62
54	2959,16	369,35	-1209,17	2172,62	1918,46	1342,03	1270,32	1364,34	1364,34	4,86
55	3212,08	622,27	252,92	-4115,57	2175,86	-751,68	-840,49	-729,37	-729,37	545,67
56	2726,11	136,30	-485,97	1283,61	1694,36	1149,91	1044,02	1172,22	1172,22	-0,52
57	1348,29	-1241,52	-1377,82	-2072,67	321,02	-286,21	-409,19	-263,90	-263,90	-1262,63
58	1817,31	-772,50	469,02	761,18	794,51	78,51	-61,57	100,82	100,82	-465,57
59	4711,7	2121,89	2894,39	1848,53	3693,38	2887,45	2730,28	2909,76	2909,76	2317,15
60	1013,85	-1575,96	-3697,85	-841,05	0,00	-128,78	-303,05	-106,47	-106,47	-2069,82
xbarra		0,00	-4,47	-137,95	1468,45	0,00	0,00	0,00	0,00	5,77
61	397,4	-2192,41	-616,45	-755,36	-611,98	-887,68	-1027,54	-809,59	-809,59	-1805,85
62	1040,13	-1549,68	26,28	-46,23	35,23	-1359,05	-1524,55	-1280,97	-1280,97	-1446,30
63	3466,98	877,17	2453,13	1594,11	2466,55	613,62	422,47	691,70	691,70	913,12
64	1479,01	-1110,80	465,16	18,03	483,06	749,96	533,17	828,04	828,04	-1090,90
65	4979,71	2389,90	3965,86	811,38	3988,23	2297,59	2055,16	2375,67	2375,67	2405,98
66	3938,02	1348,21	2924,17	978,86	2951,02	2547,13	2279,06	2625,21	2625,21	1363,38
67	10167,73	7577,92	9153,88	6955,65	9185,20	6430,20	6136,49	6508,29	6508,29	7592,87
68	1235,83	-1353,98	221,98	-1490,28	257,77	-114,13	-433,48	-36,05	-36,05	-1339,08
69	1284,98	-1304,83	271,13	-63,31	311,40	-123,28	-468,27	-45,20	-45,20	-1289,94
70	2566,82	-22,99	1552,97	749,51	1597,71	1054,26	683,62	1132,34	1132,34	-8,11
71	4321,68	1731,87	3307,83	-390,02	3357,05	2723,67	2327,39	2801,75	2801,75	1746,75
72	6412,35	3822,54	5398,50	5398,50	5452,19	5495,96	5074,04	5574,04	5574,04	3837,42

9.5.9. Total líquido por tipo de lavagem – cluster 6

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	3919,66	1944,50	-	-	-	558,79	428,06	397,12	397,12	3,92
2	6954,45	4979,29	3034,79	-	3073,32	1846,34	1712,93	1684,68	1684,68	3034,79
3	1946,39	-28,77	-5008,06	-	-1896,22	-1910,10	-2046,21	-2071,77	-2071,77	-5008,06
4	1905,18	-69,98	-41,21	-	-1898,90	-753,03	-891,83	-914,70	-914,70	-41,21
5	2648,45	673,29	743,27	-	-1117,10	-154,75	-296,24	-316,42	-316,42	743,27
6	2353,78	378,62	-294,67	-	-1373,25	-85,15	-229,33	-246,82	-246,82	-294,67
7	2497,99	522,83	144,21	-	-1190,51	-88,47	-235,34	-250,14	-250,14	144,21
8	859,22	-1115,94	-1638,77	-	-2790,75	-438,74	-588,30	-600,41	-600,41	-1638,77
9	546,73	-1428,43	-312,49	-	-3064,71	-601,47	-753,72	-763,14	-763,14	-312,49
10	2361,3	386,14	1814,57	-	-1211,62	376,02	221,08	214,36	214,36	1814,57
11	1613,77	-361,39	-747,53	-	-1920,62	-930,71	-1088,34	-1092,37	-1092,37	-747,53
12	3503,59	1528,43	1889,82	-	7,73	709,00	548,67	547,33	547,33	1889,82
13	3629,06	1653,90	125,47	-290,60	171,72	476,56	556,72	557,39	557,39	125,47
14	3117,6	1142,44	-511,46	-3836,85	-301,21	-1782,14	-1703,32	-1701,31	-1701,31	-511,46
15	4238,25	2263,09	1120,65	2291,86	857,97	590,12	667,59	670,96	670,96	1120,65
16	3271,5	1296,34	-966,75	1366,32	-70,26	821,65	897,78	902,49	902,49	-966,75
17	3340,26	1365,10	68,76	691,81	37,03	745,42	820,20	826,26	826,26	68,76
18	3186,39	1211,23	-153,87	832,61	-78,31	955,82	1029,26	1036,66	1036,66	-153,87
19	3174,79	1199,63	-11,60	676,80	-51,38	796,69	868,78	877,53	877,53	-11,60
20	676,56	-1298,60	-2498,23	-182,66	-2511,09	-413,04	-342,29	-332,20	-332,20	-2498,23
21	675,24	-1299,92	-1,32	128,51	-2473,88	-264,59	-195,19	-183,76	-183,76	-1,32
22	1821,25	-153,91	1146,01	-540,05	-1289,34	44,34	112,39	125,17	125,17	1146,01
23	3181,07	1205,91	1359,82	1567,30	109,00	844,96	911,67	925,79	925,79	1359,82
24	2072,87	97,71	-1108,20	-1430,72	-960,67	-513,36	-448,00	-432,52	-432,52	-1108,20
25	3777,42	1802,26	1704,55	148,36	782,41	994,95	1140,47	1156,62	1156,62	1704,55
26	8866,73	6891,57	5089,31	5749,13	5910,24	4337,02	4482,55	4498,69	4498,69	5089,31
27	5680,28	3705,12	-3186,45	1442,03	2762,32	2402,19	2547,71	2563,86	2563,86	-3186,45
28	1308,09	-667,07	-4372,19	-1963,41	-1571,34	-771,72	-626,20	-610,05	-610,05	-4372,19
29	1348,36	-626,80	40,27	-1991,90	-1492,55	-876,44	-730,92	-714,77	-714,77	40,27
30	725,4	-1249,76	-622,96	-2460,99	-2076,98	-1135,13	-989,61	-973,47	-973,47	-622,96
31	2113,14	137,98	1387,74	-1061,65	-650,71	105,08	250,60	266,75	266,75	1387,74
32	171	-1804,16	-1942,14	-505,56	-2554,32	-548,56	-403,04	-386,89	-386,89	-1942,14
33	98,28	-1876,88	-72,72	-576,96	-2588,52	-471,52	-325,99	-309,85	-309,85	-72,72
34	487,86	-1487,30	389,58	-1333,39	-2160,41	-919,02	-773,49	-757,35	-757,35	389,58
35	1556,42	-418,74	1068,56	-1624,65	-1053,32	-409,66	-264,13	-247,99	-247,99	1068,56
36	1435,33	-539,83	-121,09	-637,54	-1135,89	-780,86	-635,34	-619,20	-619,20	-121,09
37	1013,82	-961,34	-421,51	-2763,60	-1518,87	-1236,95	-1171,58	-1156,11	-1156,11	-421,51
38	1370,66	-604,50	356,84	-7496,07	-1123,50	-2627,34	-2560,64	-2546,51	-2546,51	356,84
39	1154,99	-820,17	-215,67	-4525,29	-1300,65	-1591,40	-1523,35	-1510,57	-1510,57	-215,67
40	1967,64	-7,52	812,65	659,55	-449,47	419,53	488,93	500,36	500,36	812,65
41	2494,78	519,62	527,14	1146,42	116,20	801,68	872,42	882,51	882,51	527,14
42	771,16	-1204,00	-1723,62	45,76	-1568,90	-557,67	-485,58	-476,84	-476,84	-1723,62
43	278,1	-1697,06	-493,06	-1835,04	-2023,43	-1198,26	-1124,82	-1117,43	-1117,43	-493,06
44	598,05	-1377,11	319,95	427,05	-1664,95	410,19	484,97	491,02	491,02	319,95
45	49,86	-1925,30	-548,19	-48,42	-2174,61	11,76	87,89	92,60	92,60	-548,19
46	787,68	-1187,48	737,82	299,82	-1398,27	-87,49	-10,02	-6,66	-6,66	737,82
47	2289,67	314,51	1501,99	733,25	142,25	855,30	934,11	936,13	936,13	1501,99
48	1614,27	-360,89	-675,40	178,94	-494,62	-70,22	9,94	10,61	10,61	-675,40
49	764,05	-1211,11	-850,22	-249,77	-1306,32	-793,35	-953,67	-955,02	-955,02	-850,22
50	1530,75	-444,41	766,70	160,09	-501,09	-1773,88	-1931,52	-1935,55	-1935,55	766,70
51	2562,21	587,05	1031,46	1407,22	568,90	509,19	354,25	347,52	347,52	1031,46
52	1138,31	-836,85	-1423,90	-829,33	-816,48	283,57	131,32	121,90	121,90	-1423,90
53	483,82	-1491,34	-654,49	-2010,96	-1432,44	-515,91	-665,47	-677,58	-677,58	-654,49
54	1457,6	-517,56	973,78	686,44	-420,13	822,14	675,27	660,47	660,47	973,78
55	1167,95	-807,21	-289,65	889,85	-671,25	384,96	240,78	223,29	223,29	-289,65
56	484,64	-1490,52	-683,31	-113,41	-1316,04	990,15	848,66	828,48	828,48	-683,31
57	670,54	-1304,62	185,90	620,68	-1091,61	1325,81	1187,02	1164,15	1164,15	185,90
58	767,95	-1207,21	97,41	-19,73	-955,67	586,15	450,04	424,48	424,48	97,41
59	381,11	-1594,05	-386,84	-1908,56	-1303,99	-359,89	-493,31	-521,56	-521,56	-386,84
60	1646,57	-328,59	1265,46	32,30	0,00	655,45	524,72	493,78	493,78	1265,46
xbarra		0,00	-38,53	-376,15	-821,65	0,00	0,00	0,00	0,00	-37,82
61	4217,65	2242,49	2571,08	3453,60	2609,61	3515,29	2983,76	2949,45	2949,45	2571,08
62	1021,02	-954,14	-625,55	-509,73	-548,50	-1428,58	-1956,07	-1994,41	-1994,41	-625,55
63	1419,09	-556,07	-227,48	-1143,12	-111,90	221,11	-302,35	-344,73	-344,73	-227,48
64	812,68	-1162,48	-833,89	-325,63	-679,78	812,98	293,55	247,14	247,14	-833,89
65	1214,46	-760,70	-432,11	730,64	-239,48	1069,77	554,38	503,93	503,93	-432,11
66	1134,94	-840,22	-511,63	-322,66	-280,47	1354,52	843,16	788,68	788,68	-511,63
67	797,02	-1178,14	-849,55	-370,93	-579,86	869,07	361,75	303,23	303,23	-849,55
68	274,04	-1701,12	-1372,53	-210,60	-1064,31	1634,59	1131,31	1068,75	1068,75	-1372,53
69	311,29	-1663,87	-1335,28	-359,25	-988,54	1821,60	1322,36	1255,76	1255,76	-1335,28
70	1266,92	-708,24	-379,65	498,97	5,62	1940,15	1444,94	1374,32	1374,32	-379,65
71	710,02	-1265,14	-936,55	328,91	-512,75	824,05	332,88	258,22	258,22	-936,55
72	510,58	-1464,58	-1135,99	-1135,99	-673,67	374,50	-112,64	-191,34	-191,34	-1135,99

9.5.10. Total líquido por tipo de lavagem – cluster 7

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	9902,15	8126,07	-	-	-	2193,19	684,18	3733,18	3733,18	9,90
2	6932,89	5156,81	-2969,26	-	-2802,78	643,10	-600,77	2183,10	2183,10	-2492,80
3	6741,63	4965,55	-191,26	-	-2827,57	857,15	-121,60	2397,14	2397,14	-1458,11
4	1903,25	127,17	-4838,38	-	-7499,47	-3225,87	-3939,49	-1685,88	-1685,88	-5600,99
5	11986,57	10210,49	10083,32	-	2750,32	5435,28	4986,79	6975,27	6975,27	6479,89
6	8615,61	6839,53	-3370,96	-	-454,16	2912,40	2729,04	4452,39	4452,39	781,38
7	13766,75	11990,67	5151,14	-	4863,45	6652,86	6734,64	8192,86	8192,86	5646,00
8	4612	2835,92	-9154,75	-	-4124,82	-447,88	-100,97	1092,12	1092,12	-5503,08
9	2565,61	789,53	-2046,39	-	-6004,74	-2181,19	-1569,16	-641,20	-641,20	-5603,81
10	2212,8	436,72	-352,81	-	-6191,07	-2194,08	-1316,92	-654,09	-654,09	-3977,85
11	3114,96	1338,88	902,16	-	-5122,44	-1700,76	-558,46	-160,77	-160,77	-1671,72
12	885,55	-890,53	-2229,41	-	-7185,37	-3434,69	-2027,26	-1894,69	-1894,69	-3311,14
13	7480,35	5704,27	6594,80	-2421,80	-424,09	3245,43	2541,72	2475,44	2475,44	4452,01
14	1748,3	-27,78	-5732,05	-5184,59	-5989,67	-1067,44	-1638,59	-1837,44	-1837,44	-2850,96
15	21,3	-1754,78	-1727,00	-6720,33	-7550,19	-2389,13	-2827,72	-3159,13	-3159,13	-3571,98
16	3702,93	1926,85	3681,63	1799,68	-3702,09	2047,86	1741,84	1277,86	1277,86	1370,04
17	267,2	-1508,88	-3435,73	-11719,37	-6971,34	-2810,05	-2983,50	-3580,04	-3580,04	-2549,11
18	41,25	-1734,83	-225,95	-8574,36	-7030,82	-2187,91	-2228,80	-2957,91	-2957,91	-1875,60
19	852,23	-923,85	810,98	-12914,52	-6053,36	-2787,61	-2695,93	-3557,61	-3557,61	-402,81
20	9,36	-1766,72	-842,87	-4602,64	-6729,76	-1576,47	-1352,23	-2346,47	-2346,47	-1103,55
21	606,55	-1169,53	597,19	-1959,06	-5966,09	-666,21	-309,40	-1436,20	-1436,20	-116,97
22	143,82	-1632,26	-462,73	-2068,98	-6262,35	-789,02	-299,64	-1559,01	-1559,01	-538,43
23	103,35	-1672,73	-40,47	-3011,61	-6136,34	-1238,32	-616,39	-2008,32	-2008,32	-388,91
24	16,15	-1759,93	-87,20	-869,40	-6057,07	-830,04	-75,54	-1600,04	-1600,04	-338,88
25	82,14	-1693,94	65,99	-7398,21	-5824,60	-2218,72	-2167,93	-3758,72	-3758,72	-153,32
26	1133,65	-642,43	1051,51	-614,65	-4606,61	251,97	302,76	-1288,03	-1288,03	952,29
27	706,43	-1069,65	-427,22	685,13	-4867,36	230,05	280,84	-1309,94	-1309,94	189,05
28	0	-1776,08	-706,43	-3702,93	-5407,31	278,98	329,77	-1261,01	-1261,01	-584,08
29	193,22	-1582,86	193,22	-73,98	-5047,62	-949,97	-899,18	-2489,97	-2489,97	-184,77
30	136,35	-1639,73	-56,87	95,10	-4938,01	-158,76	-107,97	-1698,75	-1698,75	-176,44
31	130,55	-1645,53	-5,80	-721,68	-4777,34	-1575,23	-1524,44	-3115,23	-3115,23	-119,98
32	5,1	-1770,98	-125,45	-4,26	-4736,31	353,33	404,12	-1186,67	-1186,67	-203,10
33	96,9	-1679,18	91,80	-509,65	-4478,04	758,20	808,99	-781,80	-781,80	-39,63
34	57,85	-1718,23	-39,05	-85,97	-4350,61	1059,07	1109,86	-480,93	-480,93	-64,70
35	358,49	-1417,59	300,64	255,14	-3883,50	950,87	1001,66	-589,12	-589,12	258,77
36	22,1	-1753,98	-336,39	5,95	-4053,41	1109,96	1160,75	-430,03	-430,03	-168,93
37	499,95	-1276,13	477,85	417,81	-3409,08	-1406,85	-652,35	-2176,85	-2176,85	368,53
38	1826,3	50,22	1326,35	692,65	-1916,26	1338,68	1960,61	568,68	568,68	1564,84
39	2149,15	373,07	322,85	1442,72	-1426,93	2066,83	2556,20	1296,84	1296,84	1335,53
40	190,9	-1585,18	-1958,25	190,90	-3218,71	863,94	1220,75	93,94	93,94	-1093,96
41	530,2	-1245,88	339,30	336,98	-2712,93	-218,93	5,31	-988,93	-988,93	-368,68
42	207,06	-1569,02	-323,14	70,71	-2869,60	306,01	397,69	-463,99	-463,99	-561,71
43	1082,8	-693,28	875,74	952,25	-1827,38	-228,92	-269,81	-998,92	-998,92	512,23
44	0	-1776,08	-1082,80	-5,10	-2743,71	742,29	568,83	-27,71	-27,71	-751,31
45	233,33	-1542,75	233,33	136,43	-2343,90	1288,69	982,67	518,69	518,69	-252,88
46	75,85	-1700,23	-157,48	18,00	-2334,91	1471,13	1032,54	701,13	701,13	-321,13
47	981,27	-794,81	905,42	622,78	-1263,01	1967,71	1396,56	1197,71	1197,71	697,60
48	1256,75	-519,33	275,48	1234,65	-821,06	2738,67	2034,96	1968,68	1968,68	726,93
49	1239,69	-536,39	-17,06	739,74	-671,64	-1813,05	-405,62	-273,05	-273,05	453,37
50	467,25	-1308,83	-772,44	-1359,05	-1277,60	-1166,31	-24,01	373,69	373,69	-479,04
51	463,35	-1312,73	-3,90	-1685,80	-1115,03	-764,90	112,27	775,09	775,09	-313,91
52	507,99	-1268,09	44,64	317,09	-903,91	35,10	647,13	1575,09	1575,09	-158,51
53	438,74	-1337,34	-69,25	-91,46	-806,69	-1456,33	-1109,42	83,67	83,67	-171,83
54	175,25	-1600,83	-263,49	-31,81	-903,70	-871,73	-789,96	668,26	668,26	-374,69
55	396,56	-1379,52	221,31	-686,24	-515,92	-2061,10	-2244,45	-521,10	-521,10	-21,17
56	1332,38	-443,70	935,82	1332,38	586,38	928,73	480,25	2468,73	2468,73	922,12
57	891,09	-884,99	-441,29	657,76	311,56	800,51	86,90	2340,51	2340,51	155,46
58	203,56	-1572,52	-687,53	127,71	-209,49	452,90	-525,84	1992,90	1992,90	-586,93
59	180	-1596,08	-23,56	-801,27	-66,58	20,51	-1223,37	1560,50	1560,50	-403,39
60	80,1	-1695,98	-99,90	-1176,65	0,00	416,09	-1092,92	1956,08	1956,08	-360,95
xbarra		0,00	-166,48	-1393,00	-3269,54	0,00	0,00	0,00	0,00	-406,60
61	2128,86	352,78	2048,76	889,17	2215,24	-3609,81	-1600,24	1780,18	1780,18	1815,17
62	0	-1776,08	-80,10	-467,25	252,85	-4319,49	-2707,62	1070,49	1070,49	-313,69
63	1754,57	-21,51	1674,47	1291,22	2173,90	-2159,61	-945,44	3230,37	3230,37	1440,88
64	35,7	-1740,38	-44,40	-472,29	621,50	-3123,12	-2306,64	2266,86	2266,86	-277,99
65	357,24	-1418,84	277,14	-81,50	1109,52	-4223,76	-3804,97	1166,23	1166,23	43,55
66	1471	-305,08	1390,90	1295,75	2389,75	-2261,91	-2240,83	3128,07	3128,07	1157,31
67	446,77	-1329,31	366,67	50,21	1532,00	-4696,82	-5073,43	693,16	693,16	133,08
68	550,38	-1225,70	470,28	-782,00	1802,08	-2539,20	-3313,50	2850,78	2850,78	236,69
69	936,06	-840,02	855,96	44,97	2354,24	-1840,45	-3012,44	3549,54	3549,54	622,37
70	381,48	-1394,60	301,38	177,92	1966,13	-2055,11	-3624,80	3334,88	3334,88	67,79
71	473,29	-1302,79	393,19	293,29	2224,42	-2372,13	-4339,52	3017,85	3017,85	159,60
72	843,99	-932,09	763,89	763,89	2761,60	-1505,95	-3871,04	3884,03	3884,03	530,30

9.5.11. Total líquido por tipo de lavagem – cluster 8

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	1064,17	-198,43	-	-	-	-790,89	-731,27	-945,35	-945,35	-193,35
2	1916,64	654,04	852,47	-	847,74	624,88	665,89	470,42	470,42	682,29
3	1094,87	-167,73	-821,77	-	21,24	-18,94	3,45	-173,41	-173,41	-260,52
4	568,29	-694,31	-526,58	-	-510,07	-22,90	-19,13	-177,37	-177,37	-815,65
5	1552,28	289,68	983,99	-	469,20	884,50	869,66	730,04	730,04	466,69
6	956,11	-306,49	-596,17	-	-131,70	49,52	16,06	-104,95	-104,95	-209,67
7	64,93	-1197,67	-891,18	-	-1027,61	-1163,36	-1215,43	-1317,82	-1317,82	-1206,18
8	211,8	-1050,80	146,87	-	-885,47	-148,08	-218,77	-302,54	-302,54	-738,29
9	388,26	-874,34	176,46	-	-713,74	-217,91	-307,22	-372,38	-372,38	-385,72
10	1653,33	390,73	1265,07	-	546,60	207,88	99,96	53,42	53,42	809,11
11	1333,2	70,60	-320,13	-	221,74	182,10	55,56	27,63	27,63	188,34
12	758,85	-503,75	-574,35	-	-357,34	-414,53	-559,69	-569,00	-569,00	-613,09
13	962,7	-299,90	203,85	-101,47	-158,21	-1161,85	-1089,27	-1084,62	-1084,62	-217,19
14	1386,45	123,85	423,75	-530,19	260,81	-174,80	-111,53	-97,57	-97,57	297,64
15	406,7	-855,90	-979,75	-688,17	-723,67	-976,61	-922,65	-899,38	-899,38	-816,32
16	165,57	-1097,03	-241,13	-402,72	-969,53	-695,12	-650,46	-617,88	-617,88	-954,32
17	1057,35	-205,25	891,78	-494,93	-82,48	120,08	155,43	197,31	197,31	213,76
18	1057,39	-205,21	0,04	101,28	-87,17	-118,70	-92,66	-41,46	-41,46	86,47
19	2224,73	962,13	1167,34	2159,80	1075,44	726,95	743,68	804,18	804,18	1047,38
20	957,34	-305,26	-1267,39	745,54	-196,67	327,97	335,39	405,20	405,20	-459,75
21	1723,03	460,43	765,69	1334,77	564,29	847,36	845,48	924,60	924,60	296,02
22	979,7	-282,90	-743,33	-673,63	-183,77	-735,24	-746,44	-658,01	-658,01	-311,20
23	2533,6	1271,00	1553,90	1200,40	1365,40	1113,00	1092,50	1190,23	1190,23	1218,13
24	2613,66	1351,06	80,06	1854,81	1440,73	1170,78	1140,97	1248,01	1248,01	1151,12
25	6844,53	5581,93	4230,87	5881,83	5666,87	4604,95	4647,72	4759,41	4759,41	5005,87
26	2342,88	1080,28	-4501,65	956,43	1160,49	666,60	709,37	821,06	821,06	-383,19
27	2763,65	1501,05	420,77	2356,95	1576,54	1265,31	1308,08	1419,78	1419,78	-17,22
28	840,45	-422,15	-1923,20	674,88	-351,39	-135,26	-92,50	19,20	19,20	-984,86
29	846,4	-416,20	5,95	-210,95	-350,17	-205,90	-163,13	-51,43	-51,43	-681,34
30	695,89	-566,71	-150,51	-361,50	-505,41	-595,22	-552,46	-440,76	-440,76	-387,92
31	2468,46	1205,86	1772,57	243,73	1262,43	855,65	898,42	1010,12	1010,12	1414,18
32	1189,98	-72,62	-1278,48	232,64	-20,78	445,58	488,35	600,05	600,05	-193,49
33	439,93	-822,67	-750,05	-1283,10	-775,56	-550,76	-508,00	-396,30	-396,30	-1091,27
34	1262,66	0,06	822,73	282,96	42,44	-567,31	-524,54	-412,84	-412,84	181,42
35	413,25	-849,35	-849,41	-2120,35	-811,69	-1122,38	-1079,61	-967,91	-967,91	-663,34
36	531,57	-731,03	118,32	-2082,09	-698,10	-1026,34	-983,57	-871,87	-871,87	-561,01
37	513,09	-749,51	-18,48	-6331,44	-721,31	-1687,05	-1716,87	-1609,82	-1609,82	-407,19
38	486,51	-776,09	-26,58	-1856,37	-752,62	-1150,33	-1170,84	-1073,10	-1073,10	-457,36
39	1152,71	-109,89	666,20	-1610,94	-91,15	-306,19	-317,39	-228,96	-228,96	218,58
40	705,35	-557,25	-447,36	-135,10	-543,24	-230,93	-232,82	-153,70	-153,70	-359,44
41	886,89	-375,71	181,54	40,49	-366,43	-125,97	-118,55	-48,74	-48,74	-240,23
42	710,51	-552,09	-176,38	14,62	-547,53	-541,17	-524,44	-463,94	-463,94	-350,35
43	587,33	-675,27	-123,18	-1881,13	-675,44	-986,04	-960,01	-908,81	-908,81	-479,32
44	239,62	-1022,98	-347,71	-950,36	-1027,88	-465,34	-430,00	-388,11	-388,11	-760,91
45	443,68	-818,92	204,06	3,75	-828,55	-507,58	-462,92	-430,35	-430,35	-456,93
46	3134,92	1872,32	2691,24	1872,26	1857,96	1344,39	1398,35	1421,62	1421,62	2272,88
47	610,53	-652,07	-2524,39	197,28	-671,16	-885,66	-822,39	-808,43	-808,43	-851,43
48	1769,97	507,37	1159,44	1238,40	483,55	251,50	324,08	328,73	328,73	203,51
49	1041,09	-221,51	-728,88	528,00	-250,05	-965,15	-1110,31	-1119,62	-1119,62	-179,18
50	1476,6	214,00	435,51	990,09	180,73	33,66	-92,88	-120,81	-120,81	137,65
51	1301,43	38,83	-175,17	148,72	0,83	36,43	-71,49	-118,03	-118,03	41,75
52	1826,59	563,99	525,16	1121,24	521,26	1084,21	994,90	929,75	929,75	502,08
53	146,25	-1116,35	-1680,34	-740,64	-1163,81	-672,71	-743,40	-827,18	-827,18	-1245,56
54	2263,35	1000,75	2117,10	1552,84	948,56	1205,57	1153,50	1051,11	1051,11	1095,07
55	1946,27	683,67	-317,08	1358,94	626,75	566,80	533,34	412,33	412,33	732,14
56	350,93	-911,67	-1595,34	111,31	-973,32	-160,13	-174,97	-314,60	-314,60	-1288,11
57	1186,25	-76,35	835,32	742,57	-142,72	428,89	432,67	274,43	274,43	-51,77
58	1346,91	84,31	160,66	-1788,01	13,21	-249,72	-227,33	-404,19	-404,19	306,61
59	2015,23	752,63	668,32	1404,70	676,80	712,94	753,95	558,48	558,48	748,59
60	1343,16	80,56	-672,07	-426,81	0,00	18,59	78,21	-135,87	-135,87	-97,92
xbarra		0,00	4,73	97,53	59,93	0,00	0,00	0,00	0,00	-0,89
61	807,08	-455,52	-536,08	-234,01	-540,81	-850,80	-1154,07	-1391,42	-1391,42	-650,67
62	363,72	-898,88	-979,44	-1112,88	-988,90	-730,86	-1006,20	-1271,48	-1271,48	-962,00
63	857,77	-404,83	-485,39	-443,66	-499,58	-58,86	-306,28	-599,49	-599,49	-467,36
64	554,01	-708,59	-789,15	-1272,58	-808,06	160,00	-59,50	-380,63	-380,63	-740,44
65	294,96	-967,64	-1048,20	148,71	-1071,84	-175,64	-367,21	-716,26	-716,26	-993,06
66	1629,98	367,38	286,82	-633,37	258,45	920,57	756,92	379,94	379,94	350,39
67	1349,21	86,61	6,05	-597,06	-27,05	318,10	182,38	-222,52	-222,52	72,84
68	1211,37	-51,23	-131,79	860,44	-169,62	1048,67	940,88	508,05	508,05	-62,39
69	2485,17	1222,57	1142,01	1298,92	1099,45	2076,18	1996,31	1535,55	1535,55	1212,69
70	1973,46	710,86	630,30	626,55	583,01	725,19	673,25	184,57	184,57	701,85
71	1431,65	169,05	88,49	-583,58	36,47	477,73	453,70	-62,90	-62,90	160,51
72	1175,63	-86,97	-167,53	-167,53	-224,27	199,43	203,33	-341,20	-341,20	-95,21

9.5.12. Total líquido por tipo de peça – cluster 1

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	255,1109524	-115,41	-	-	-	-64,02	-73,96	-31,97	-31,97	0,26
2	365,2769048	-5,24	110,17	-	107,55	-5,04	-11,32	27,02	27,02	89,78
3	257,7295238	-112,79	-107,55	-	-2,62	-141,35	-143,99	-109,30	-109,30	-51,33
4	141,2659524	-229,26	-116,46	-	-121,71	-89,02	-88,00	-56,97	-56,97	-144,98
5	243,2761905	-127,25	102,01	-	-22,32	-61,27	-56,61	-29,22	-29,22	2,46
6	334,0233333	-36,50	90,75	-	65,81	-8,08	0,24	23,98	23,98	91,70
7	428,9721429	58,45	94,95	-	158,14	-27,80	-15,83	4,26	4,26	158,94
8	142,5419048	-227,98	-286,43	-	-130,91	-18,81	-3,19	13,25	13,25	-173,52
9	237,6257143	-132,90	95,08	-	-38,45	-35,01	-15,74	-2,96	-2,96	-28,01
10	341,9678571	-28,55	104,34	-	63,27	55,45	78,38	87,51	87,51	84,40
11	409,7940476	39,27	67,83	-	128,48	53,27	79,84	85,32	85,32	127,77
12	433,2057143	62,68	23,41	-	149,27	122,98	153,21	155,04	155,04	114,23
13	409,4245238	38,90	-23,78	154,31	122,87	95,82	80,70	79,79	79,79	57,42
14	398,6595238	28,14	-10,76	33,38	109,48	33,87	20,58	17,84	17,84	30,06
15	355,6554762	-14,87	-43,00	97,93	63,86	-37,91	-49,37	-53,93	-53,93	-21,63
16	301,512381	-69,01	-54,14	160,25	7,09	76,75	67,11	60,72	60,72	-69,52
17	418,5345238	48,01	117,02	175,26	121,50	119,50	111,69	103,48	103,48	67,59
18	400,6357143	30,11	-17,90	66,61	100,98	64,06	58,07	48,03	48,03	30,16
19	564,757619	194,24	164,12	135,79	262,48	113,51	109,35	97,48	97,48	185,57
20	167,825	-202,70	-396,93	25,28	-137,07	12,00	9,66	-4,03	-4,03	-264,98
21	310,347381	-60,17	142,52	72,72	2,83	43,23	42,73	27,21	27,21	-45,90
22	325,597381	-44,92	15,25	-16,37	15,46	44,60	45,92	28,58	28,58	-17,38
23	342,8595238	-27,66	17,26	-66,93	30,10	-8,15	-5,00	-24,17	-24,17	4,90
24	299,5245238	-71,00	-43,34	-133,68	-15,86	-5,18	-0,21	-21,20	-21,20	-39,85
25	382,2588095	11,74	82,73	-27,17	64,26	42,12	31,97	10,06	10,06	54,40
26	553,47	182,95	171,21	154,81	232,85	162,14	152,00	130,09	130,09	209,89
27	715,0235714	344,50	161,55	359,37	391,78	294,93	284,78	262,87	262,87	310,80
28	180,5290476	-189,99	-534,49	-120,98	-145,33	-70,77	-80,91	-102,82	-102,82	-313,50
29	222,5759524	-147,95	42,05	-195,96	-105,91	-102,99	-113,13	-135,04	-135,04	-180,87
30	265,3807143	-105,14	42,80	-135,26	-65,72	-97,73	-107,87	-129,78	-129,78	-85,80
31	447,8295238	77,31	182,45	-116,93	114,11	-29,95	-40,10	-62,00	-62,00	121,44
32	242,1788095	-128,34	-205,65	74,35	-94,17	59,82	49,67	27,76	27,76	-119,30
33	126,3604762	-244,16	-115,82	-183,99	-212,60	-167,29	-177,43	-199,34	-199,34	-200,65
34	192,2788095	-178,24	65,92	-133,32	-149,31	-115,25	-125,39	-147,30	-147,30	-76,75
35	209,7228571	-160,80	17,44	-133,14	-134,48	-167,82	-177,96	-199,87	-199,87	-37,13
36	179,8738095	-190,65	-29,85	-119,65	-166,95	-151,36	-161,51	-183,41	-183,41	-56,25
37	242,9119048	-127,61	63,04	-139,35	-106,54	-155,81	-150,85	-171,84	-171,84	23,04
38	279,4938095	-91,03	36,58	-273,98	-72,57	-170,42	-167,27	-186,44	-186,44	52,96
39	271,4545238	-99,07	-8,04	-443,57	-83,23	-207,23	-205,91	-223,25	-223,25	29,62
40	372,5119048	1,99	101,06	191,98	15,20	62,63	62,12	46,60	46,60	122,12
41	373,4309524	2,91	0,92	150,86	13,50	-10,72	-13,05	-26,74	-26,74	87,75
42	438,5983333	68,08	65,17	173,22	76,05	16,90	12,74	0,88	0,88	127,57
43	476,4319048	105,91	37,83	28,60	111,26	-59,93	-65,92	-75,96	-75,96	128,54
44	248,3392857	-122,18	-228,09	6,16	-119,45	7,39	-0,42	-8,63	-8,63	-136,69
45	520,7521429	150,23	272,41	394,39	150,34	168,52	158,88	152,49	152,49	175,22
46	471,1302381	100,61	-49,62	278,85	98,10	105,02	93,56	88,99	88,99	74,97
47	482,2335714	111,71	11,10	272,51	106,58	46,11	32,82	30,08	30,08	64,41
48	494,117619	123,60	11,88	314,24	115,85	104,30	89,18	88,27	88,27	57,68
49	571,2647619	200,74	77,15	328,35	190,37	81,90	112,13	113,95	113,95	118,16
50	519,9961905	149,47	-51,27	240,50	136,48	-20,55	6,02	11,50	11,50	32,75
51	660,8807143	290,36	140,88	389,43	274,75	91,56	114,49	123,61	123,61	164,17
52	420,9321429	50,41	-239,95	48,42	32,18	20,41	39,68	52,46	52,46	-123,21
53	530,2585714	159,74	109,33	156,83	138,88	55,47	71,09	87,53	87,53	21,72
54	537,18	166,66	6,92	98,58	143,19	24,85	36,82	56,90	56,90	22,36
55	631,1752381	260,65	94,00	154,74	234,56	4,17	12,49	36,22	36,22	109,90
56	271,1869048	-99,33	-359,99	22,85	-128,05	-60,40	-55,73	-28,34	-28,34	-281,85
57	433,4116667	62,89	162,22	-87,34	31,56	-9,46	-8,44	22,59	22,59	-38,18
58	366,9259524	-3,60	-66,49	-104,20	-37,55	-89,82	-92,46	-57,77	-57,77	-93,64
59	603,3488095	232,83	236,42	121,12	196,25	76,59	70,30	108,64	108,64	169,84
60	409,717619	39,20	-193,63	-84,40	0,00	-70,74	-80,68	-38,69	-38,69	-72,86
xbarra		0,00	2,62	49,28	38,75	0,00	0,00	0,00	0,00	10,85
61	496,3328571	125,81	86,62	-74,93	83,99	-115,72	-50,09	-3,54	-3,54	34,80
62	267,2580952	-103,26	-142,46	-252,74	-147,70	-395,98	-335,83	-283,80	-283,80	-194,27
63	492,005	121,48	82,29	-168,88	74,43	-200,00	-145,33	-87,82	-87,82	30,48
64	336,4109524	-34,11	-73,31	-84,52	-83,79	-186,80	-137,60	-74,62	-74,62	-125,12
65	350,95	-19,57	-58,77	-179,31	-71,87	-246,53	-202,81	-134,34	-134,34	-110,58
66	393,35	22,83	-16,37	-143,83	-32,09	-241,67	-203,43	-129,49	-129,49	-68,18
67	603,7488095	233,23	194,03	-27,43	175,69	-145,95	-113,18	-33,76	-33,76	142,22
68	337,4171429	-33,10	-72,30	66,23	-93,26	-116,86	-89,57	-4,67	-4,67	-124,11
69	605,7992857	235,28	196,08	172,39	172,50	40,24	62,05	152,42	152,42	144,27
70	535,1097619	164,59	125,39	168,18	99,19	-44,33	-28,00	67,85	67,85	73,58
71	460,227619	89,71	50,51	-143,12	21,69	-189,23	-178,37	-77,04	-77,04	-1,30
72	359,3942857	-11,13	-50,32	-50,32	-81,77	-243,75	-238,37	-131,57	-131,57	-102,13

9.5.13. Total líquido por tipo de peça – cluster 2

Período	Observado	Resíduos (€)								
		Mean	Naive	SNaive	Drift	Modelo 1	Modelo 2	Modelo 3	Modelo 4	ARIMA
1	20235,86	5775,46	-	-	-	2455,21	-570,62	6885,47	6885,47	20,24
2	18200,14	3739,74	-2035,72	-	-1988,95	209,39	-2168,09	4639,65	4639,65	-1361,10
3	13850,69	-609,71	-4349,45	-	-6291,62	-2028,56	-3757,68	2401,70	2401,70	-3890,73
4	11174,42	-3285,98	-2676,27	-	-8921,12	-2842,39	-3923,15	1587,87	1587,87	-4199,88
5	26996,26	12535,86	15821,84	-	6947,50	7441,07	7008,66	11871,33	11871,33	9787,74
6	18533,93	4073,53	-8462,33	-	-1468,06	2474,86	2690,81	6905,12	6905,12	-851,73
7	29185,69	14725,29	10651,76	-	9230,47	9342,20	10206,51	13772,46	13772,46	7750,42
8	10074,64	-4385,76	-19111,05	-	-9833,80	730,27	2242,93	5160,53	5160,53	-10269,96
9	9406,57	-5053,83	-668,07	-	-10455,10	-2853,01	-692,00	1577,25	1577,25	-6527,50
10	17028,01	2567,61	7621,44	-	-2786,88	-364,17	2445,20	4066,08	4066,08	2084,00
11	14669,36	208,96	-2358,65	-	-5098,76	-6052,32	-2594,59	-1622,06	-1622,06	-607,51
12	9721,57	-4738,83	-4947,79	-	-9999,78	-6821,05	-2714,97	-2390,79	-2390,79	-4188,82
13	13763,2	-697,20	4041,63	-6472,66	-5911,37	1407,01	-646,03	-808,12	-808,12	-287,52
14	9681,89	-4778,51	-4081,31	-8518,25	-9945,91	-2884,40	-4613,27	-5099,53	-5099,53	-2793,02
15	7182,28	-7278,12	-2499,61	-6668,41	-12398,75	-3272,51	-4677,19	-5487,64	-5487,64	-1284,95
16	10964,34	-3496,06	3782,06	-210,08	-8569,91	2372,00	1291,49	156,87	156,87	4708,79
17	14477,98	17,58	3513,64	-12518,28	-5009,50	347,25	-409,08	-1867,88	-1867,88	-3741,85
18	8342,1	-6118,30	-6135,88	-10191,83	-11098,60	-2292,51	-2724,66	-4507,64	-4507,64	-2969,74
19	16526,86	2066,46	8184,76	-12658,83	-2867,07	2107,84	1999,86	-107,29	-107,29	-287,34
20	3922,55	-10537,85	-12604,31	-6152,09	-15424,61	2,64	218,84	-2212,49	-2212,49	-688,83
21	7215,04	-7245,36	3292,49	-2191,53	-12085,34	379,92	920,30	-1835,21	-1835,21	3311,65
22	13180,46	-1279,94	5965,42	-3847,55	-6073,15	1212,74	2077,30	-1002,39	-1002,39	3084,88
23	11943,82	-2516,58	-1236,64	-2725,54	-7263,01	-3353,40	-2164,66	-5568,53	-5568,53	2060,38
24	12307,04	-2153,36	363,22	2585,47	-6853,02	1188,88	2701,80	-1026,25	-1026,25	4697,64
25	15320,01	859,61	3012,97	1556,81	-3793,28	3958,03	3417,90	-472,23	-472,23	3207,23
26	10942,71	-3517,69	-4377,30	1260,82	-8123,80	-629,38	-1169,51	-5059,64	-5059,64	80,56
27	10713,61	-3746,79	-229,10	3531,33	-8306,13	1253,03	712,90	-3177,23	-3177,23	1398,52
28	6441,11	-8019,29	-4272,50	-4523,23	-12531,85	-1157,03	-1697,16	-5587,29	-5587,29	-5845,22
29	8427,78	-6032,62	1986,67	-6050,20	-10498,41	-4708,74	-5248,87	-9139,00	-9139,00	-3656,39
30	8235,85	-6224,55	-191,93	-106,25	-10643,57	-1404,56	-1944,69	-5834,82	-5834,82	1547,45
31	11909,95	-2550,45	3674,10	-4616,91	-6922,69	-1514,87	-2055,00	-5945,13	-5945,13	-595,20
32	3651,79	-10808,61	-8258,16	-270,76	-15134,08	726,08	185,95	-3704,18	-3704,18	-637,06
33	5985,8	-8474,60	2334,01	-1229,24	-12753,30	144,88	-395,24	-4285,38	-4285,38	-120,67
34	9233,82	-5226,58	3248,02	-3946,64	-9458,50	-1739,70	-2279,83	-6169,96	-6169,96	-594,45
35	14608,46	148,06	5374,64	2664,64	-4037,09	305,45	-234,68	-4124,81	-4124,81	5808,59
36	13095,4	-1365,00	-1513,06	788,36	-5503,37	2971,45	2431,32	-1458,81	-1458,81	1657,29
37	12530,5	-1929,90	-564,90	-2789,51	-6021,50	-2267,54	-754,63	-4482,67	-4482,67	-1499,87
38	15418,11	957,71	2887,61	4475,40	-3087,12	409,96	1598,70	-1805,17	-1805,17	4777,37
39	12633,72	-1826,68	-2784,39	1920,11	-5824,73	-262,92	601,64	-2478,05	-2478,05	156,85
40	8046,91	-6413,49	-4586,81	1605,80	-10364,77	-2987,29	-2446,91	-5202,42	-5202,42	-1793,57
41	13212,22	-1248,18	5165,31	4784,44	-5152,68	-3360,36	-3144,16	-5575,49	-5575,49	2859,32
42	13647,62	-812,78	435,40	5411,77	-4670,51	571,15	463,18	-1643,98	-1643,98	2230,43
43	13935,92	-524,48	288,30	2025,97	-4335,44	-2924,96	-3357,11	-5140,09	-5140,09	-729,21
44	7284,87	-7175,53	-6651,05	3633,08	-10939,71	923,10	166,77	-1292,03	-1292,03	-1856,39
45	7077,99	-7382,41	-206,88	1092,19	-11099,82	-2198,98	-3279,49	-4414,11	-4414,11	-2769,26
46	14774,58	314,18	7696,59	5540,76	-3356,45	365,00	-1039,68	-1850,13	-1850,13	4021,97
47	28039,44	13579,04	13264,86	13430,98	9955,18	10300,37	8571,51	8085,24	8085,24	12220,79
48	20170,86	5710,46	-7868,58	7075,46	2133,37	6610,85	4557,81	4395,72	4395,72	241,42
49	17111,65	2651,25	-3059,21	4581,15	-879,06	-5552,71	-1446,63	-1122,45	-1122,45	-2560,72
50	25768,9	11308,50	8657,25	10350,79	7824,96	2894,44	6352,16	7324,69	7324,69	5332,57
51	25073,92	10613,52	-694,98	12440,20	7176,75	4310,96	7120,33	8741,22	8741,22	4186,91
52	23515,23	9054,83	-1558,69	15468,32	5664,84	4614,71	6775,73	9044,97	9044,97	3792,24
53	24719,68	10259,28	1204,45	11507,46	6916,06	280,78	1793,44	4711,04	4711,04	158,26
54	21593,84	7133,44	-3125,84	7946,22	3837,00	651,06	1515,36	5081,32	5081,32	-3308,50
55	17716,99	3256,59	-3876,85	3781,07	6,92	-7010,21	-6794,26	-2579,95	-2579,95	-5995,73
56	11845,99	-2614,41	-5871,00	4561,12	-5817,31	-2382,09	-2814,50	2048,17	2048,17	-5175,20
57	21670,48	7210,08	9824,49	14592,49	4053,96	4527,19	3446,43	8957,45	8957,45	6926,04
58	22802,03	8341,63	1131,55	8027,45	5232,28	526,14	-1202,98	4956,40	4956,40	319,22
59	24405,29	9944,89	1603,26	-3634,15	6882,32	-1200,10	-3577,57	3230,16	3230,16	-6597,03
60	17476,2	3015,80	-6929,09	-2694,66	0,00	-3950,13	-6975,96	480,13	480,13	-5815,43
xbarra		0,00	-46,77	1137,98	-4470,13	0,00	0,00	0,00	0,00	82,14
61	18682,23	4221,83	1206,03	1570,58	1252,80	-16278,70	-9039,33	-772,80	-772,80	-263,69
62	18761,97	4301,57	1285,77	-7006,93	1379,32	-16409,07	-10142,23	-903,16	-903,16	-5657,78
63	19419,99	4959,59	1943,79	-5653,93	2084,11	-13639,54	-8345,24	1866,36	1866,36	-4560,34
64	12426,04	-2034,36	-5050,16	-11089,19	-4863,06	-18771,05	-14449,28	-3265,14	-3265,14	-10568,75
65	12824,06	-1636,34	-4652,14	-11895,62	-4418,27	-23911,41	-20562,17	-8405,51	-8405,51	-10932,29
66	18217,05	3756,65	740,85	-3376,79	1021,49	-15022,31	-12645,60	483,60	483,60	-3562,88
67	22236,99	7776,59	4760,79	4520,00	5088,21	-14786,78	-13382,61	719,13	719,13	2908,32
68	5936,87	-8523,53	-11539,33	-5909,12	-11165,14	-20587,79	-20156,15	-5081,88	-5081,88	-9679,66
69	7677,46	-6782,94	-9798,74	-13993,02	-9377,77	-21762,41	-22303,30	-6256,50	-6256,50	-14150,93
70	18699,57	4239,17	1223,37	-4102,46	1691,11	-15872,90	-17386,32	-366,99	-366,99	-3844,28
71	19546,78	5086,38	2070,58	-4858,51	2585,09	-18355,18	-20841,14	-2849,28	-2849,28	-4010,78
72	20563,96	6103,56	3087,76	3087,76	3649,05	-13158,94	-16617,43	2346,96	2346,96	1387,54